

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2005, 10:27:07 ; Search time 652 Seconds  
(without alignments)  
8734.342 Million cell updates/sec

Title: US-10-800-161-24

Perfect score: 962

Sequence: 1 tgrgtttctcagaataagca.....aacattttcttctggttaac 962

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq\_16Dec04.\*

- 1: Geneseqn1980s.\*
- 2: Geneseqn1990s.\*
- 3: Geneseqn2000s.\*
- 4: Geneseqn2001as.\*
- 5: Geneseqn2001bs.\*
- 6: Geneseqn2002as.\*
- 7: Geneseqn2002bs.\*
- 8: Geneseqn2003as.\*
- 9: Geneseqn2003bs.\*
- 10: Geneseqn2003cs.\*
- 11: Geneseqn2003ds.\*
- 12: Geneseqn2004as.\*
- 13: Geneseqn2004bs.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	962	100.0	962	ADRI16843	Adri16843 Arabidops
2	862	89.6	862	ADRI16844	Adri16844 Arabidops
3	862	89.6	1700	ADRI17259	Adri17259 Arabidops
4	862	89.6	1700	ADRI16822	Adri16822 Arabidops
5	544	56.5	544	ADRI16846	Adri16846 Arabidops
6	274	28.5	274	ADRI16845	Adri16845 Arabidops
7	272.4	28.3	274	ADRI16847	Adri16847 Arabidops
8	86	8.9	537	ACN53339	ACN53339 Cotton an
9	76.8	8.0	7657	AAS45477	Aas45477 Chemicall
10	76.8	8.0	7657	ABL34022	Abl34022 Human imm
11	76.8	8.0	14006	ABL33958	Abl33958 Human imm
12	76.6	8.0	529	ACN56029	ACN56029 Cotton an
13	76.4	7.9	6644	AA333181	Aax33181 Base sequ
14	76.4	7.9	7372	AA333182	Aax33182 Base sequ
15	76.4	7.9	7797	AA333180	Aax33180 Cowpox vi
16	76.4	7.9	7996	AA333184	Aax33184 Base sequ
17	76	7.9	19124	AA333184	Aax33184 Base sequ
18	76	7.9	19124	AA333184	Aax33184 Base sequ
19	75.6	7.9	700	ACD92384	ACD92384 Human col
20	74.8	7.8	9539	AAS45347	Aas45347 Chemicall

C	21	74.8	7.8	9539	6	ABK28180	Abk28180 DNA trans
	22	74.4	7.7	627	13	ACN54555	ACN54555 Cotton an
	23	74.4	7.7	2520	12	ADJ48201	Adj48201 Maize oil
	24	74.2	7.7	517	13	ACN56273	ACN56273 Cotton an
	25	73.8	7.7	529	8	ABX52180	ABX52180 Bovine ES
	26	73.4	7.6	113515	6	ABL34174	Abl34174 Human imm
	27	73.2	7.6	478	13	ACN51993	ACN51993 Cotton an
	28	72.6	7.5	16033	6	ABL33404	Abl33404 Human imm
	29	72.2	7.5	778	6	ABQ15588	Abq15588 Oligonucle
	30	72.2	7.5	778	6	ABQ15589	Abq15589 Oligonucle
	31	72	7.5	446	8	ABX38235	ABX38235 Bovine ES
	32	72	7.5	16167	6	ABL33083	Abl33083 Human imm
	33	72	7.5	16167	6	ABL34529	Abl34529 Human met
	34	72	7.5	16167	6	ABL70254	AbL70254 Chemicall
	35	72	7.5	16167	7	ADL99790	AdL99790 Complemen
	36	71.8	7.5	34571	13	ADR66967	ADR66967 Mouse can
	37	71.6	7.4	17934	6	ABL33719	Abl33719 Human imm
	38	71.6	7.4	110000	10	ABQ84281	ABQ84281 Continuation (2 of
	39	71.6	7.4	110000	10	ABQ84281	ABQ84281 Continuation (3 of
	40	71.4	7.4	3505	12	ADJ48200	Adj48200 Maize oil
	41	71.2	7.4	10620	2	AAX02996	Aax02996 Human IL-
	42	71.2	7.4	14690	2	AAX22303	Aax22303 Human IL-
	43	71	7.4	447	13	ACN54792	ACN54792 Cotton an
	44	71	7.4	15548	6	ABL34155	Abl34155 Human imm
	45	70.8	7.4	3683	8	ABZ10199	Abz10199 Haematopo

#### ALIGNMENTS

##### RESULT 1

ADRI16843  
ID ADRI16843 standard; DNA; 962 BP.

AC ADRI16843;

XX 21-OCT-2004 (first entry)

XX Arabidopsis thaliana N16 promoter DNA #1.

XX N16; transgenic; herbicide resistance; mouse-ear cross; promoter; ds.

XX Arabidopsis thaliana.

XX US2004154051-A1.

XX 05-AUG-2004.

XX 12-MAR-2004; 2004US-00800161.

XX 15-DEC-1999; 99US-0171008P.

XX 11-JAN-2000; 2000US-0175519P.

XX 08-DEC-2000; 2000US-0073368S.

XX 20-JAN-2004; 2004US-00760752.

(SYGN ) SYNGENTA PARTICIPATIONS AG.

XX Cade RM, Dietrich RA, Lawton KA;

XX WPI; 2004-580223/56.

XX Novel isolated nucleic acid molecule comprising Arabidopsis N16 promoter, useful in regulating transcription of coding sequence of interest.

XX Claim 11; SEQ ID NO 24; 39pp; English.

XX The present invention relates to an Arabidopsis N16 gene promoter useful in regulating transcription of coding sequence of interest. The invention is useful in production of transgenic plant or seed exhibiting resistance to herbicides and microorganism such as Phytophthora parasitica, Pseudomonas syringae, Cercospora nicotianae, Peronospora parasitica. The present sequence is Arabidopsis thaliana N16 promoter DNA. This sequence



```
Db 241 AGAAGAGAGAACATGTAGTGAACAAATAAAGAGATATGATGATATATTTATGAGAG 300
Qy 401 GTGCTGAAGATTATTTTAGGAGAGGAGAGAAATAGAAAAGAAATGACATGGTGAA 460
Db 301 GTGCTGAAGATTATTTTAGGAGAGGAGAGAAATAGAAAAGAAATGACATGGTGAA 360
Qy 461 TCTGAAGAGATGAATTTGTGTTAAAGATGAAGAGAGAAAGAACTCCATGGCTTAAAGTC 520
Db 361 TCTGAAGAGATGAATTTGTGTTAAAGATGAAGAGAGAAAGAACTCCATGGCTTAAAGTC 420
Qy 521 TCGTAAAGAGATCAAAAGAGAAACAAAGAGAGGAGAAAGAAAGGCTTAAATAGA 580
Db 421 TCGTAAAGAGATCAAAAGAGAAACAAAGAGAGGAGAAAGAAAGGCTTAAATAGA 480
Qy 581 CTAACATATGCCAAAATTTCTGTAGCCGACAAATACTATTTGGTCCAAAGGTTATTTGTG 640
Db 481 CTAACATATGCCAAAATTTCTGTAGCCGACAAATACTATTTGGTCCAAAGGTTATTTGTG 540
Qy 641 TATTCCTTTGAAGTCAAAAGTTATTTCTTACATATACTCTTAAATAATATAGCCGATACAA 700
Db 541 TATTCCTTTGAAGTCAAAAGTTATTTCTTACATATACTCTTAAATAATATAGCCGATACAA 600
Qy 701 TTTTCCACACATCGATCTCTTTATTTCCAAAAGTCAATAAGTGTGACGTCATGATACT 760
Db 601 TTTTCCACACATCGATCTCTTTATTTCCAAAAGTCAATAAGTGTGACGTCATGATACT 660
Qy 761 TAGCCTTTAAACATCGCATGATGATGTCAATAGCATCAATCTCCACCGTCCCAATTTATT 820
Db 661 TAGCCTTTAAACATCGCATGATGATGTCAATAGCATCAATCTCCACCGTCCCAATTTATT 720
Qy 821 TAGTTGTTGACATATCGACCGTCTAAGTTCCACACCGCGCTTAAAGAGTTTCATTAT 880
Db 721 TAGTTGTTGACATATCGACCGTCTAAGTTCCACACCGCGCTTAAAGAGTTTCATTAT 780
Qy 881 AAATTTTACCAATAAATCAGCAATAATTTTCTTGACTAAGCTTAAACGACCCG 940
Db 781 AAATTTTACCAATAAATCAGCAATAATTTTCTTGACTAAGCTTAAACGACCCG 840
Qy 941 TTAACATTTTCTTCTGGCTAAC 962
Db 841 TTAACATTTTCTTCTGGCTAAC 862
```

## RESULT 3

```
ADN17259
ID ADN17259 standard; DNA; 1700 BP.
XX
AC ADN17259;
XX
XX 17-JUN-2004 (first entry)
XX Arabidopsis thaliana N116 gene.
XX N116; SAR; systemic acquired resistance; mouse-ear cress; gene; db.
XX Arabidopsis thaliana.
XX
XX Key Location/Qualifiers
FT misc_feature 365..374
FT /*tag= a
FT /*note= "TC1 motif"
FT misc_feature 426..435
FT /*tag= b
FT /*note= "TC1 motif"
FT misc_feature 609..614
FT /*tag= c
FT /*note= "MYCATR22 element"
FT misc_feature 646..665
FT /*tag= d
FT /*note= "CMV AS1 salicylic acid response element"
FT misc_feature 707..712
FT /*tag= e
FT /*note= "PAL box"
```

```
FT misc_feature 757..762
FT /*tag= f
FT /*note= "HEXAMERAT 4 element"
FT CDS 863..1231
FT /*tag= g
FT /*product= "Arabidopsis thaliana N116 protein"
XX
PN US6706952-B1.
XX
PD 16-MAR-2004.
XX
XX 08-DEC-2000; 2000US-00733685.
PF
XX 15-DEC-1999; 99US-0171008P.
PR 11-JAN-2000; 2000US-0175519P.
XX
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
PA
XX Cad RM, Dietrich RA;
PI
XX WPI; 2004-313378/29.
XX P-PSDB; ADN17258.
DR
XX New N116 nucleic acid sequence and encoded protein, useful for increasing
PT systemic acquired resistance gene expression in a plant.
XX
XX Example 2; SEQ ID NO 3; 29pp; English.
XX
XX The invention relates to Arabidopsis N116 gene encoding a protein
CC involved in the regulation of SAR gene expression in plants. The N116
CC nucleic acid molecule and the encoded protein is useful in increasing
CC systemic acquired resistance (SAR) gene expression in a plant. The
CC present sequence is Arabidopsis thaliana N116 gene.
XX
SQ Sequence 1700 BP; 593 A; 237 C; 361 G; 509 T; 0 U; 0 Other;
Query Match 89.6%; Score 862; DB 12; Length 1700;
Best Local Similarity 100.0%; Pred. No. 2.1e-146;
Matches 862; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 101 TGGGTTTTTATTCGATAAACAATGACAAATATTTATTTATTTTCATGAGTTTATTTGGATAG 160
Db 1 TGGGTTTTTATTCGATAAACAATGACAAATATTTATTTATTTTCATGAGTTTATTTGGATAG 60
Qy 161 CATGACAAATATTTAATATATATCAGTGTAAATAACATGTTTGTCTTAAATAACATGCAAT 220
Db 61 CATGACAAATATTTAATATATATCAGTGTAAATAACATGTTTGTCTTAAATAACATGCAAT 120
Qy 221 TTAATAATCAGACATTTGTTTTTAAATACTAATCTCTTATATCAACACGACATTTGAC 280
Db 121 TTAATAATCAGACATTTGTTTTTAAATACTAATCTCTTATATCAACACGACATTTGAC 180
Qy 281 GGAATAATTCAGGTAAAGAGAGAAATAAAGAAATGAGAGATAGAGAGATTTCTATGGAAAA 340
Db 181 GGAATAATTCAGGTAAAGAGAGAGAAATAAAGAAATGAGAGATAGAGAGATTTCTATGGAAAA 240
Qy 341 AGAAGAGAGACATGCTAGGTGAACAAATAAAGAGATATGATGATATATTTTATGAGAG 400
Db 241 AGAAGAGAGACATGCTAGGTGAACAAATAAAGAGATATGATGATATATTTTATGAGAG 300
Qy 401 GTGGTGAAGATTTATTTTAGGAGAGGAGAGAGAAATAGAAAAGAAATACACATGGTGAA 460
Db 301 GTGGTGAAGATTTATTTTAGGAGAGGAGAGAGAGAAATAGAAAAGAAATACACATGGTGAA 360
Qy 461 TCTGAAAGAGATGAATTTGTGTTAAAGATGAAGAGAGAAAGAGAACTCCATGGCTTAAAGTC 520
Db 361 TCTGAAAGAGATGAATTTGTGTTAAAGATGAAGAGAGAAAGAGAACTCCATGGCTTAAAGTC 420
Qy 521 TCGTAAAGAGATGAAGAGAGAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 580
Db 421 TCGTAAAGAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Qy 581 CTAACATTTTCCCAAAATTTCTTGAGCCGACAAATACTATTTGGTCCAAAGGTTATTTGTG 640
```





QY 761 TAGCCTTTAAACAATCCGATGATGTCATTAGCATCAATCTCCACCGTCCCAATTTATT 820  
DB 661 TAGCCTTTAAACAATCCGATGATGTCATTAGCATCAATCTCCACCGTCCCAATTTATT 720  
QY 821 TAGTTGTTGACAAATATCGACGCTTAAGTTCCACACGCGGCTATAAGAGTTTCATTAT 880  
DB 721 TAGTTGTTGACAAATATCGACGCTTAAGTTCCACACGCGGCTATAAGAGTTTCATTAT 780  
QY 881 AAATTTTGTAGCAAAATAAATAACAGCAAAATAATTTTCTTGACTAAGCTTAAACGACGCGG 940  
DB 781 AAATTTTGTAGCAAAATAAATAACAGCAAAATAATTTTCTTGACTAAGCTTAAACGACGCGG 840  
QY 941 TTAACATTTTCTTGCGCTAAC 962  
DB 841 TTAACATTTTCTTGCGCTAAC 862

RESULT 5  
ADRI6846  
ID ADR16846 standard; DNA; 544 BP.  
XX AC  
XX ADR16846;  
DT 21-OCT-2004 (first entry)  
XX DE  
XX Arabidopsis thaliana Nil6 promoter DNA #4.  
XX KW  
XX Nil6; transgenic; herbicide resistance; mouse-ear cress; promoter; ds.  
XX OS  
XX Arabidopsis thaliana.  
XX PN  
XX US2004154051-A1.  
XX PD  
XX 05-AUG-2004.  
XX PF  
XX 12-MAR-2004; 2004US-00800161.  
XX PR  
XX 15-DEC-1999; 99US-0171008P.  
XX PR  
XX 11-JAN-2000; 2000US-0175519P.  
XX PR  
XX 08-DEC-2000; 2000US-00733685.  
XX PR  
XX 20-JAN-2004; 2004US-00760752.  
XX XX  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX PA  
XX Cade RM, Dietrich RA, Lawton KA;  
XX PI  
XX WPI; 2004-580223/56.  
XX DR  
XX Novel isolated nucleic acid molecule comprising Arabidopsis Nil6  
XX PT promoter, useful in regulating transcription of coding sequence of  
XX PT interest.  
XX XX  
XX Claim 12; SEQ ID NO 27; 39pp; English.

CC The present invention relates to an Arabidopsis Nil6 gene promoter useful  
CC in regulating transcription of coding sequence of interest. The invention  
CC is useful in production of transgenic plant or seed exhibiting resistance  
CC to herbicides and microorganism such as Phytophthora parasitica,  
CC Pseudomonas syringae, Cercospora nicotianae, Peronospora parasitica. The  
CC present sequence is Arabidopsis thaliana Nil6 promoter DNA. This sequence  
CC is used in the exemplification of the invention.  
XX XX  
SQ Sequence 544 BP; 209 A; 84 C; 95 G; 156 T; 0 U; 0 Other;

Query Match 56.5%; Score 544; DB 13; Length 544;  
Best Local Similarity 100.0%; Pred. No. 4.5e-89;  
Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 408 AGATTATTTTACGAGAGGGAGAGAAATAGAAAAAATGACATGTTGAATCTGAAG 467  
DB 1 AGATTATTTTACGAGAGGGAGAGAAATAGAAAAAATGACATGTTGAATCTGAAG 60

QY 468 AAGATGAATTTGTGTTAAAGATGAAGAGAGAAAGAACTCCATGGCTAAAGTCTCGTAAA 527  
DB 61 AAGATGAATTTGTGTTAAAGATGAAGAGAGAAAGAACTCCATGGCTAAAGTCTCGTAAA 120  
QY 528 GAAGATGAAAAAGAAACAAAGAGAGGAGAGAGAAAGAGAAAGGCTAAAAATAGACTAACTA 587  
DB 121 GAAGATGAAAAAGAAACAAAGAGAGGAGAGAGAAAGAGAAAGGCTAAAAATAGACTAACTA 180  
QY 588 TTGCCAAAAATTTCTGTAGCCGACAAATATCTATTGGTCCAAAGGTTATTTTGTGTATTTCTT 647  
DB 181 TTGCCAAAAATTTCTGTAGCCGACAAATATCTATTGGTCCAAAGGTTATTTTGTGTATTTCTT 240  
QY 648 TTGAAGTCAAAAGTTATTTCTTACATATCTCTAAAAATATAGCCGATACCAATTTTTC 707  
DB 241 TTGAAGTCAAAAGTTATTTCTTACATATCTCTAAAAATATAGCCGATACCAATTTTTC 300  
QY 708 ACACATGGGACTTCTTTTATTTCCAAAAGTCAATTAAGTGTGACGTCATGATACCTTACGCTT 767  
DB 301 ACACATGGGACTTCTTTTATTTCCAAAAGTCAATTAAGTGTGACGTCATGATACCTTACGCTT 360  
QY 768 TAAAAACATCGCATGATGATGTCATTAGCATCAATCTCCACGCTCCAAATTTTATTAGTTGT 827  
DB 361 TAAAAACATCGCATGATGATGTCATTAGCATCAATCTCCACGCTCCAAATTTTATTAGTTGT 420  
QY 828 TGAACAATATCGACCGTCTAAGTTCCACACGCGGCTATAAGAGTTTTCATTATAAAATTTT 887  
DB 421 TGAACAATATCGACCGTCTAAGTTCCACACGCGGCTATAAGAGTTTTCATTATAAAATTTT 480  
QY 888 AGCAAAATAAATCAGCAAAATAATTTTCTTGACTAAGCTTAAACGACGCGTTAAACAT 947  
DB 481 AGCAAAATAAATCAGCAAAATAATTTTCTTGACTAAGCTTAAACGACGCGTTAAACAT 540  
QY 948 TTTC 951  
DB 541 TTTC 544

RESULT 6  
ADRI6845  
ID ADR16845 standard; DNA; 274 BP.  
XX AC  
XX ADR16845;  
DT 21-OCT-2004 (first entry)  
XX DE  
XX Arabidopsis thaliana Nil6 promoter DNA #3.  
XX KW  
XX Nil6; transgenic; herbicide resistance; mouse-ear cress; promoter; ds.  
XX OS  
XX Arabidopsis thaliana.  
XX PN  
XX US2004154051-A1.  
XX PD  
XX 05-AUG-2004.  
XX PF  
XX 12-MAR-2004; 2004US-00800161.  
XX PR  
XX 15-DEC-1999; 99US-0171008P.  
XX PR  
XX 11-JAN-2000; 2000US-0175519P.  
XX PR  
XX 08-DEC-2000; 2000US-00733685.  
XX PR  
XX 20-JAN-2004; 2004US-00760752.  
XX XX  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX PA  
XX Cade RM, Dietrich RA, Lawton KA;  
XX PI  
XX WPI; 2004-580223/56.  
XX DR  
XX Novel isolated nucleic acid molecule comprising Arabidopsis Nil6  
XX PT promoter, useful in regulating transcription of coding sequence of  
XX PT interest.  
XX XX  
XX Claim 1; SEQ ID NO 26; 39pp; English.

XX The present invention relates to an Arabidopsis N116 gene promoter useful  
 CC in regulating transcription of coding sequence of interest. The invention  
 CC is useful in production of transgenic plant or seed exhibiting resistance  
 CC to herbicides and microorganism such as Phytophthora parasitica,  
 CC Pseudomonas syringae, Cercospora nicotianae, Peronospora parasitica. The  
 CC present sequence is Arabidopsis thaliana N116 promoter DNA. This sequence  
 CC is used in the exemplification of the invention.  
 XX SQ Sequence 274 BP; .92 A; 58 C; 35 G; 89 T; 0 U; 0 Other;

Query Match 28.5%; Score 274; DB 13; Length 274;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-40;  
 Matches 274; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 678 TCTAAAAATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTCCTAAAAAGTCA 737  
 DB 1 TCTAAAAATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTCCTAAAAAGTCA 60  
 QY 738 ATAAAGTGTGACGTGATGATCTTACGCTTTAAACATCGCATGATGTCATTAGCAT 797  
 DB 61 ATAAAGTGTGACGTGATGATCTTACGCTTTAAACATCGCATGATGTCATTAGCAT 120  
 QY 798 CAATCTCCACCGTCCAAATTTATTAGTTGTGACAATATCGACCGCTAAAGTTCCACACC 857  
 DB 121 CAATCTCCACCGTCCAAATTTATTAGTTGTGACAATATCGACCGCTAAAGTTCCACACC 180  
 QY 858 GACGGCTATAAGAGTTTCATTATAAAATTTTAGCAAAATAAAATCAGCAAAATAATTTTTC 917  
 DB 181 GACGGCTATAAGAGTTTCATTATAAAATTTTAGCAAAATAAAATCAGCAAAATAATTTTTC 240  
 QY 918 TTGACTAAGCTTAAACGACCGCTTAACATTTTC 951  
 DB 241 TTGACTAAGCTTAAACGACCGCTTAACATTTTC 274

RESULT 7  
 ADR16847  
 ID ADR16847 standard; DNA; 274 BP.  
 XX AC ADR16847;  
 XX DT 21-OCT-2004 (first entry)  
 XX DE Arabidopsis thaliana N116 promoter DNA #5.  
 XX KW N116; transgenic; herbicide resistance; mouse-ear cross; promoter; ds.  
 XX OS Arabidopsis thaliana.  
 XX PN US2004154051-A1.  
 XX PD 05-AUG-2004.  
 XX PF 12-MAR-2004; 2004US-00800161.  
 XX PR 15-DEC-1999; 99US-0171008P.  
 XX PR 11-JAN-2000; 2000US-0175519P.  
 XX PR 08-DEC-2000; 2000US-00733685.  
 XX PR 20-JAN-2004; 2004US-00760752.  
 XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX PA Cade RM, Dietrich RA, Lawton KA;  
 XX PI WPI; 2004-580223/56.  
 XX DR  
 XX PT Novel isolated nucleic acid molecule comprising Arabidopsis N116  
 PT promoter, useful in regulating transcription of coding sequence of  
 PT interest.  
 XX PS Claim 1; SEQ ID NO 28; 39pp; English.  
 XX

CC The present invention relates to an Arabidopsis N116 gene promoter useful  
 CC in regulating transcription of coding sequence of interest. The invention  
 CC is useful in production of transgenic plant or seed exhibiting resistance  
 CC to herbicides and microorganism such as Phytophthora parasitica,  
 CC Pseudomonas syringae, Cercospora nicotianae, Peronospora parasitica. The  
 CC present sequence is Arabidopsis thaliana N116 promoter DNA. This sequence  
 CC is used in the exemplification of the invention.  
 XX SQ Sequence 274 BP; 91 A; 58 C; 36 G; 89 T; 0 U; 0 Other;

Query Match 28.3%; Score 272.4; DB 13; Length 274;  
 Best Local Similarity 99.6%; Pred. No. 4.3e-40;  
 Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 678 TCTAAAAATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTCCTAAAAAGTCA 737  
 DB 1 TCTAGAAATATATAGCCGATACCAATTTTCCACACATGGACTTCCTTTATTCCTAAAAAGTCA 60  
 QY 738 ATAAAGTGTGACGTGATGATCTTACGCTTTAAACATCGCATGATGTCATTAGCAT 797  
 DB 61 ATAAAGTGTGACGTGATGATCTTACGCTTTAAACATCGCATGATGTCATTAGCAT 120  
 QY 798 CAATCTCCACCGTCCAAATTTATTAGTTGTGACAATATCGACCGCTAAAGTTCCACACC 857  
 DB 121 CAATCTCCACCGTCCAAATTTATTAGTTGTGACAATATCGACCGCTAAAGTTCCACACC 180  
 QY 858 GACGGCTATAAGAGTTTCATTATAAAATTTTAGCAAAATAAAATCAGCAAAATAATTTTTC 917  
 DB 181 GACGGCTATAAGAGTTTCATTATAAAATTTTAGCAAAATAAAATCAGCAAAATAATTTTTC 240  
 QY 918 TTGACTAAGCTTAAACGACCGCTTAACATTTTC 951  
 DB 241 TTGACTAAGCTTAAACGACCGCTTAACATTTTC 274

RESULT 8  
 ACN53339  
 ID ACN53339 standard; cDNA; 537 BP.  
 XX AC ACN53339;  
 XX DT 02-DEC-2004 (first entry)  
 XX DE Cotton androecium tissue EST Clone ID: LIB3828-006-Q1-N6-D9, SEQ:8120.  
 XX KW Cotton; plant; EST; expressed sequence tag; transgenic plant; androecium;  
 KW variety Nutcoton33B; library LIB3828; molecular tag; molecular marker;  
 KW genetic mapping; molecular mapping; seed germination; plant growth;  
 KW plant quality; plant yield; plant breeding; tissue printing; ss.  
 XX OS Gossypium hirsutum.  
 XX PN US2004123340-A1.  
 XX PD 24-JUN-2004.  
 XX PF 12-DEC-2001; 2001US-00021323.  
 XX PR 14-DEC-2000; 2000US-0255619P.  
 XX (DEIK/) DEIKMAN J.  
 XX (FENG/) FENG P C C.  
 XX (FINC/) FINCHER K L.  
 XX (ZIEG/) ZIEGLER T E.  
 XX Deikman J, Feng PCC, Fincher KL, Ziegler TE;  
 XX WPI; 2004-479808/45.  
 XX DR New isolated nucleic acid molecule that encodes a plant protein or its  
 PT fragment, useful for isolating a variety of agronomically significant  
 PT genes associated with plant growth, quality or yield, and as molecular  
 PT tags to map genes.















CC therapy for destroying cancer cells selectively, the treatment of  
CC autoimmune diseases and graft rejection reaction, and apoptosis induction  
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have  
CC encountered the problem where if an adenovirus vector capable of  
CC expressing an apoptosis-associated gene is introduced into animal cells,  
CC the cells producing the virus will be destroyed because the period of  
CC time required to induce cell death by apoptosis is shorter than that  
CC required to replicate and produce the virus, resulting in failure to  
CC obtain a recombinant virus having the integrated apoptosis-associated  
CC gene. In this invention an apoptosis-resistant 293 cell line (having an  
CC apoptosis resistant gene introduced) is established and overcomes the  
CC problem. The present sequence represents the cowpox virus bsr gene which  
CC is used in an example from the present invention  
XX  
SQ

Sequence 7797 BP; 2542 A; 1760 C; 1656 G; 1839 T; 0 U; 0 Other;

Query Match 7.9%; Score 76.4; DB 2; Length 7797;

Best Local Similarity 45.2%; Pred. No. 0.00012;

Matches 281; Conservative 0; Mismatches 341; Indels 0; Gaps 0;

Qy	11	AGAAATGACGAGAAATTTATTAAGCATGCAATCTCTTAGATCGCGAAGTTTAA	70
Db	4975	AA	5034
Qy	71	AAAAACATAGATGTTTACAATATTACATGGGTTTTATTGGATAACATGACAAATAT	130
Db	5035	AA	5094
Qy	131	TTATTTATTTTCATGAGTTTTTATGGATGATGACAAATATTATATATATCAGTGTAAAT	190
Db	5095	AA	5154
Qy	191	AACATGTTTGGTCTTAAATATACATGCAATTTTAAATCAGACATTTGTTTTAAATCAAA	250
Db	5155	AA	5214
Qy	251	TCTAATCTCTTATATCACACGACATTTGACGGAAATTCAGGTAAGAGAGAAATAAAG	310
Db	5215	AA	5274
Qy	311	AATGAGAGATGAGAGATTTCTATGGAAAAAGAGAGACATGTAGGTGAACAAAAT	370
Db	5275	AA	5334
Qy	371	AAAGAGATATGATATATTTTATGAGAGGTGCTGAAGATTATTTTAGGAGAGGAGAG	430
Db	5335	AA	5394
Qy	431	AGAAATAGAAAAAGAAATGACATGCTGAAAGATGAAATGCTTAAAGATGA	490
Db	5395	AA	5454
Qy	491	AGAGAGAAAGAGAACTCCATGGCTAAAGTCTCGTAAAGAGAGTGAAGAAACAAAAGA	550
Db	5455	AA	5514
Qy	551	AGGAAGAGAAAGAGAAAGGCTAAATAGACTAATCTTGCCTAAATTTCTGTAGCCGAC	610
Db	5515	AA	5574
Qy	611	AAATACTATTGGTCCAAGTTT	632
Db	5575	AAAAAGAAATTTGGTCAAGCTT	5596

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2005, 07:25:35 ; Search time 391.3 Seconds  
(without alignments)  
9357.993 Million cell updates/sec

Title: US-10-800-161-24

Perfect score: 962

Sequence: 1 tgtgtttctcagaatagca.....aacattttcttctgctgaac 962

Scoring table: IDENTITY NUC  
Gapop 10<sup>-7</sup>, Gapext 1.0

Searched: 34239544 segs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: gb\_est1:\*
- 2: gb\_est2:\*
- 3: gb\_hc:\*
- 4: gb\_est3:\*
- 5: gb\_est4:\*
- 6: gb\_est5:\*
- 7: gb\_est6:\*
- 8: gb\_gss1:\*
- 9: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	231	24.0	267	8	BH854168 SALK_0788
2	222	23.1	228	8	BZ352867 SALK_0827
3	152	15.8	152	8	BZ770600 SALK_1435
4	149.6	15.6	222	9	CC797914 SALK_1455
5	147	15.3	327	8	CC455857 SALK_0894
6	144.6	15.0	203	8	BZ381166 SALK_1163
7	128.4	13.3	387	8	BH749275 SALK_0475
8	117.4	12.2	223	8	BZ359042 SALK_1337
9	110.6	11.5	237	8	BH814587 SALK_0666
10	101.2	10.5	147	8	BZ352865 SALK_0827
11	99.8	10.4	1101	9	CNS0021J Drosophil
12	99.6	10.4	1434	8	BZ576813 msh2 5115
13	98.2	10.2	860	7	CF289324 AGENCOURT
14	97.2	10.1	1201	9	CNS0167M Drosophil
15	96	10.0	660	8	BH183498 023_L_07-
16	96	10.0	660	9	CNS070NJ T3_end of
17	95.4	9.9	1267	9	AG346514 Mus muscu
18	95	9.9	1235	9	AG361148 Mus muscu
19	94	9.8	737	9	CL077131 CH216-143
20	93.2	9.7	764	7	CF289381 AGENCOURT
21	92.6	9.6	982	8	AQ325799 nbxb0021B
22	92.6	9.6	1228	8	CU104752 ISBI-4311
23	92.4	9.6	1061	9	CU145135 ISBI-145G
24	92.4	9.6	1241	9	AG448181 Mus muscu

25	92.4	9.6	1599	9	CL083840	CL083840 ISBI-2H14
26	92.4	9.6	1872	8	BZ558518	BZ558518 pa98401.2
27	92	9.6	1101	9	CNS0039G	AL063921 Drosophil
28	91.8	9.5	1047	8	CC266341	CC266341 CH261-12K
29	91.2	9.5	842	9	CNS010QN	AL099305 Drosophil
30	91.2	9.5	1409	9	AG346237	AG346237 Mus muscu
31	91	9.5	826	7	CV483860	CV483860 AGENCOURT
32	90.8	9.4	795	7	CF288665	CF288665 AGENCOURT
33	90.4	9.4	1036	9	CNS00599	AL057797 Drosophil
34	90.4	9.4	1355	9	AG346348	AG346348 Mus muscu
35	90.2	9.4	934	8	AZ184244	AZ184244 SP_1002 B
36	89.8	9.3	1025	9	CL113815	CL113815 ISBI-59A2
37	89.6	9.3	997	9	CL112076	CL112076 ISBI-56F1
38	89.6	9.3	1059	9	CNS0022B	AL097133 Drosophil
39	89.6	9.3	1110	9	CL037168	CL037168 CH216-42N
40	89.4	9.3	755	7	CV491618	CV491618 AGENCOURT
41	89.4	9.3	1074	8	BZ696936	BZ696936 SP_Ba009
42	89.4	9.3	1210	9	CG749728	CG749728 P044-1-C0
43	89.4	9.3	1251	9	AG332167	AG332167 Mus muscu
44	89.4	9.3	1522	9	CL128484	CL128484 ISBI-9411
45	89.2	9.3	1614	9	AG365768	AG365768 Mus muscu

#### ALIGNMENTS

BH854168 267 bp DNA linear GSS 13-JUN-2002  
SALK\_078835.55.00.x Arabidopsis thaliana TDNA insertion lines  
Arabidopsis thaliana genomic clone SALK\_078835.55.00.x, genomic survey sequence.

ACCESSION BH854168

VERSION BH854168.1 GI:21425039

KEYWORDS GSS.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 267)  
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,  
Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,  
Shinn,P., Zimmerman,J. and Ecker,J.R.

AUTHORS

TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome

JOURNAL Unpublished (2001)

COMMENT Contact: Joseph R. Ecker

Salk Institute Genomic Analysis Laboratory (SIGnAL)  
The Salk Institute for Biological Studies  
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752

Fax: 858 558 6379

Email: ecker@salk.edu

This is single pass sequence recovered from the left border of TDNA.

Class: TDNA tagged.

Location/Qualifiers

1. .267

/organism="Arabidopsis thaliana"

/mol\_type="genomic DNA"

/ecotype="Col-0"

/db\_xref="taxon:3702"

/clone="SALK\_078835.55.00.x"

/clone\_lib="Arabidopsis thaliana TDNA insertion lines"

/note="PCR was performed on Arabidopsis thaliana lines

each of which contains one or more TDNA insertion

elements. The resultant fragment for each line was

directly sequenced to determine the genomic sequence at

the site of insertion. Details of the protocols used can

be found at [http://signal.salk.edu/tdna\\_protocols.html](http://signal.salk.edu/tdna_protocols.html)

ORIGIN

Query Match 24.0%; Score 231; DB 8; Length 267;  
 Best Local Similarity 94.0%; Pred. No. 2.6e-32;  
 Matches 251; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

QY 23 AATATTTTATAAAGCATGCAATTCCTTATAGATCGGAAGTTTAA-AAAAACATATA 81  
 |||||  
 DB 1 ATATATTTTATAAAGCATGCAATTCCTTATAGATCGGAAGTTTAAAAACATATA 60  
 |||||

QY 82 GAATTTGTTACAAATATTACATGGGTTTTTTATTGGATAACATGACAAATATTATTATTTTC 141  
 |||||  
 DB 61 GAATTTGTTACAAATATTACATGAGTTTTTTATTGGATAACATGACAAATATTATATATTAC 120  
 |||||

QY 142 ATGAGTTTTTTATGGATAGCATGCAAAATATTATATATATATATATATGAGTTTAAATTAACATGTTTTG 201  
 |||||  
 DB 121 ATGAAATTTTGTGGATAGCATGCAAAATATTATATATATATATATATGAGTTTAAATTAACATGTTTTG 180  
 |||||

QY 202 TTTCTTAAATACATGCATTTTAAATCAGACATTTGTTTTTAAATCAAAATCTAATCTCTTT 261  
 |||||  
 DB 181 TTTCTTAAATACATGCATTTTAAATCAGACATTTGTTTTTAAATCAAAATCTAATCTCTTT 240  
 |||||

QY 262 ATATCAACAGCATTCGACGAAAATT 288  
 |||||  
 DB 241 ATATCAACAGCATTCGATGAAGAATT 267  
 |||||

RESULT 2  
 BZ352867  
 LOCUS  
 DEFINITION  
 BZ352867 228 bp DNA linear GSS 14-NOV-2002  
 SALK\_082724.45.05.x Arabidopsis thaliana TDNA insertion lines  
 Arabidopsis thaliana genomic clone SALK\_082724.45.05.x, genomic  
 survey sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Arabidopsis thaliana (thale cress)  
 Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE  
 AUTHORS  
 Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,  
 Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,  
 Shinn,P., Zimmerman,J. and Ecker,J.R.  
 A Sequence-Indexed Library of Insertion Mutations in the  
 Arabidopsis Genome  
 Unpublished (2001)  
 Contact: Joseph R. Ecker  
 Salk Institute Genomic Analysis Laboratory (SIGNAL)  
 The Salk Institute for Biological Studies  
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
 Tel: 858 453 4100 x1752  
 Fax: 858 558 6379  
 Email: ecker@salk.edu

JOURNAL  
 COMMENT  
 This is single pass sequence recovered from the left border of  
 TDNA. This sequence lies within 300 bases of the 5' end of  
 AT3g25882.  
 Class: TDNA tagged.

FEATURES  
 source  
 Location/Qualifiers  
 1..228  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /ecotype="Col-0"  
 /db\_xref="taxon:3702"  
 /clone="SALK\_082724.45.05.x"  
 /clone\_lib="Arabidopsis thaliana TDNA insertion lines"  
 /note="PCR was performed on Arabidopsis thaliana lines  
 each of which contains one or more TDNA insertion  
 elements. The resultant fragment for each line was  
 directly sequenced to determine the genomic sequence at  
 the site of insertion. Details of the protocols used can  
 be found at [http://signal.salk.edu/tdna\\_protocols.html](http://signal.salk.edu/tdna_protocols.html)

ORIGIN

Query Match 23.1%; Score 222; DB 8; Length 228;  
 Best Local Similarity 97.4%; Pred. No. 1.2e-30;  
 Matches 222; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 701 TTTTTCACACATGGACTTCCTTTATTCCAAAAAGTCAATAAAGTGTGACGTGATGATACT 760  
 |||||  
 DB 1 TTTTTCNCNCNTGGACTTCCTTTATTCNAAAGTCNATANAGTGTGACGTGATGATACT 60  
 |||||

QY 761 TAGCCTTTAAACATCGCATGATGATGTCTATTAGCATCAATCTCCACCGTCCCAATTTATT 820  
 |||||  
 DB 61 TAGCCTTTAAACATCGCATGATGATGTCTATTAGCATCAATCTCCACCGTCCCAATTTATT 120  
 |||||

QY 821 TAGTTGTTGACAATATCGACCGTCTTAAGTTCCACACCGAGCGGTATAAGAGTTTCATTAT 880  
 |||||  
 DB 121 TAGTTGTTGACAATATCGACCGTCTTAAGTTCCACACCGAGCGGTATAAGAGTTTCATTAT 180  
 |||||

QY 881 AATATTTTACAAATAAATCAGCAATAATTTTTTCTTGACTAAAGCT 928  
 |||||  
 DB 181 AATATTTTACAAATAAATCAGCAATAATTTTTTCTTGACTAAAGCT 228  
 |||||

RESULT 3  
 BZ770600  
 LOCUS  
 DEFINITION  
 BZ770600 152 bp DNA linear GSS 13-MAR-2003  
 SALK\_143535.56.00.x Arabidopsis thaliana TDNA insertion lines  
 Arabidopsis thaliana genomic clone SALK\_143535.56.00.x, genomic  
 survey sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Arabidopsis thaliana (thale cress)  
 Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE  
 AUTHORS  
 Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,  
 Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,  
 Shinn,P., Zimmerman,J. and Ecker,J.R.  
 A Sequence-Indexed Library of Insertion Mutations in the  
 Arabidopsis Genome  
 Unpublished (2001)  
 Contact: Joseph R. Ecker  
 Salk Institute Genomic Analysis Laboratory (SIGNAL)  
 The Salk Institute for Biological Studies  
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
 Tel: 858 453 4100 x1752  
 Fax: 858 558 6379  
 Email: ecker@salk.edu

JOURNAL  
 COMMENT  
 This is single pass sequence recovered from the left border of  
 TDNA. This sequence lies within 300 bases of the 5' end of  
 AT3g25882.  
 Class: TDNA tagged.

FEATURES  
 source  
 Location/Qualifiers  
 1..152  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /ecotype="Col-0"  
 /db\_xref="taxon:3702"  
 /clone="SALK\_143535.56.00.x"  
 /clone\_lib="Arabidopsis thaliana TDNA insertion lines"  
 /note="PCR was performed on Arabidopsis thaliana lines  
 each of which contains one or more TDNA insertion  
 elements. The resultant fragment for each line was  
 directly sequenced to determine the genomic sequence at  
 the site of insertion. Details of the protocols used can  
 be found at [http://signal.salk.edu/tdna\\_protocols.html](http://signal.salk.edu/tdna_protocols.html)

ORIGIN

Query Match 15.8%; Score 152; DB 8; Length 152;  
 Best Local Similarity 100.0%; Pred. No. 7.6e-18;  
 Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 777 GCATGATGTCATTAGCATCAATCTCCACCGTCAATTTATTTAGTTGTTGACAATAT 836
Db 1 GCATGATGTCATTAGCATCAATCTCCACCGTCAATTTATTTAGTTGTTGACAATAT 60

QY 837 CGACCGTCTAAGTCTCCACCGCCTATAGAGTTTCATTATAAATTTAGCAAAATA 896
Db 61 CGACCGTCTAAGTCTCCACCGCCTATAGAGTTTCATTATAAATTTAGCAAAATA 120

QY 897 AAATCAGCAAAATAATTTTCTTGACTAAGCT 928
Db 121 AAATCAGCAAAATAATTTTCTTGACTAAGCT 152

RESULT 4
CC797914/c 222 bp DNA linear GSS 01-JUL-2003
LOCUS SALK_145557.51.25.x Arabidopsis thaliana TDNA insertion lines
DEFINITION Arabidopsis thaliana genomic clone SALK_145557.51.25.x, genomic
survey sequence.
ACCESSION CC797914.1 GI:32393137
VERSION CC797914.1
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 222)
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
Shinn,P., Zimmerman,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
Unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGnAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA.
Class: TDNA tagged.
Location/Qualifiers
1..222
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Col-0"
/db_xref="taxon:3702"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN
Query Match 15.6%; Score 149.6; DB 9; Length 222;
Best Local Similarity 85.3%; Pred. No. 2.1e-17;
Matches 191; Conservative 0; Mismatches 26; Indels 7; Gaps 2;

QY 285 AATTCAGGTAAAGAGAAAATAAGAAATGAGATGAGAGATTTCTATGGAAGAA 344
Db 222 AATTCAGGTAAAGAGAAAATAAGAAATGAGATGAGAGATTTCTATGGAAGAA 165

QY 345 AGAGAGAACATGTAGTGCACAAATAAAGAGATATGATGATATTTATGAGAGTGG 404
Db 164 AGAGAGAACATGTAGTGCACAAATAAAGAGATATGATGATATTTATGAGAGTGG 105

QY 405 TGAAGATTATTT-----TAGGAGAGGAGAGAGAAATAGAAAAAGAAAATGACATGTTGA 459

```

```

Db 104 TGAAGATATTTCTAGGAGAGGAGAGAGAAATAGAAATAGAAAAAGAAATGACATGTTGA 45
QY 460 ATCTGAAGAAGATGAATTTGTTTAAAGATGAAGAGAGAGAAAGAGA 503
Db 44 ATCTGAAGAAGATGAATTTGTTTAAAGAGAGAGAGAGAAAGAGA 1

RESULT 5
CC455857 327 bp DNA linear GSS 30-MAY-2003
LOCUS SALK_089493.54.85.x Arabidopsis thaliana TDNA insertion lines
DEFINITION Arabidopsis thaliana genomic clone SALK_089493.54.85.x, genomic
survey sequence.
ACCESSION CC455857
VERSION CC455857.1 GI:31216228
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 327)
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
Shinn,P., Zimmerman,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
Unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGnAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA.
Class: TDNA tagged.
Location/Qualifiers
1..327
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Col-0"
/db_xref="taxon:3702"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN
Query Match 15.3%; Score 147; DB 8; Length 327;
Best Local Similarity 93.7%; Pred. No. 6.1e-17;
Matches 164; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 1 TGTGTTTCTCAGAAATAGCAGAAATATTTATAAAAAAGCATGCAATTTCTTTATAGATCG 60
Db 153 TGTGTTTCTCAGAAATAGCAGATATATTTATAAAAAAGCATGCAATTTCTTTATAAATCG 212

QY 61 CGAAGT-TTAAAAAACATATAGAAATTTGTACAAATTTTACATGCGGTTTTTTATGGATAAC 119
Db 213 CGAAGTCTTTAAGAAACATATAGAAATTTGTACAAATTTACATCGGTTTTTTATGGATAAC 272

QY 120 ATGACAAATATTTATTTATTTTCATGAGTTTTTTTATGATGATGATGACAAATATTTA 174
Db 273 ATGACTAATATTTATATATATTTTCATGAGTTTTTTTGTGATGATGACAAATATTTA 327

RESULT 6

```

ACCESSION	BH749275
VERSION	BH749275.1
KEYWORDS	GI:189631963
SOURCE	GSS.
ORGANISM	Arabidopsis thaliana (thale cress)
REFERENCE	Arabis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi. 1 (bases 1 to 387) Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadzinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R. A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome Unpublished (2001) Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379 Email: ecker@salk.edu This is single pass sequence recovered from the left border of TDNA. Class: TDNA tagged. Location/Qualifiers 1..387 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /ecotype="Col-0" /db_xref="taxon:3702" /clone="SALK_047543.32.95.x" /clone_lib="Arabidopsis thaliana TDNA insertion lines" /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"
ORIGIN	
Query Match	13.3%; Score 128.4; DB 8; Length 387;
Best Local Similarity	89.8%; Pred. No. 1.5e-13;
Matches 149; Conservative	0; Mismatches 16; Indels 1; Gaps 1;
Qy	285 AATTCAGGTGAAGAGAAATAAAGATGAGATGAGATGAGAGATTCTTATGGAAAAAGAA 344
Dd	
Qy	344 AATTCAGGTGAAGAGAAATAAAGATGAGATGAGATGAGAGATTCTTATGGAAAAATGAA 285
Dd	
Qy	345 AGAGAGAACAATGAGTGTAAC-AAATAAAGAGATATGATGATATATTTTATGAGAGGTG 403
Dd	
Qy	404 GTGAAGATTATTTTAGGAGAGGAGAGAGAAATAGAAAAGAAAAT 449
Dd	
Qy	224 GTGAAGATTATCATGAGCAGGCTCGGTGCATCGAATCAGCAATT 179
Dd	
RESULT 8	
BZ359042/c	223 bp DNA linear GSS 14-NOV-2008
LOCUS	BZ359042
DEFINITION	SALK_133702.17.85.n Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_133702.17.85.n, genomic survey sequence.
ACCRESSON	BZ359042
VERSION	BZ359042.1
KEYWORDS	GI:24951589
SOURCE	GSS.
ORGANISM	Arabidopsis thaliana (thale cress)
REFERENCE	Arabis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi. 1 (bases 1 to 223)

## AUTHORS

Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R.,  
Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L.,  
Shinn, P., Zimmerman, J. and Ecker, J.R.  
A Sequence-Indexed Library of Insertion Mutations in the  
Arabidopsis Genome  
Unpublished (2001)

## TITLE

JOURNAL  
COMMENT

Contact: Joseph R. Ecker  
Salk Institute Genomic Analysis Laboratory (SIGNAL)  
The Salk Institute for Biological Studies  
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
Tel: 858 453 4100 x1752  
Fax: 858 558 6379  
Email: ecker@salk.edu

This is single pass sequence recovered from the left border of  
TDNA.

FEATURES  
source

Class: TDNA tagged.  
Location/Qualifiers  
1..223  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/ecotype="Col-0"  
/db\_xref="taxon:3702"  
/clone="SALK\_133702.17.85.n"  
/clone\_lib="Arabidopsis thaliana TDNA insertion lines"  
/note="PCR was performed on Arabidopsis thaliana lines  
each of which contains one or more TDNA insertion  
elements. The resultant fragment for each line was  
directly sequenced to determine the genomic sequence at  
the site of insertion. Details of the protocols used can  
be found at [http://signal.salk.edu/tdna\\_protocols.html](http://signal.salk.edu/tdna_protocols.html)"

## ORIGIN

Query Match 12.2%; Score 117.4; DB 8; Length 223;  
Best Local Similarity 76.7%; Pred. No. 1.6e-11;  
Matches 171; Conservative 0; Mismatches 46; Indels 6; Gaps 2;  
  
Qy 285 AATTCAGTAAAGAGAGAAAATAAAGATGAGATAGAGATTTCTATGGAAAAGAA 344  
|||||  
Db 223 AATTCAGCAAAACGGAAGCAATGCATGCGACATAGAGCTATTTCTATCCAAATAC 164  
|||||  
Qy 345 AGACAGAA-CATGTAGTGAACAAATAAAGAGATATGATATATATTTATGAGGTG 403  
|||||  
Db 163 AGAGTGGATCATGTGGTTCAACAACATACAGAGATATGATGCTATATTTCTGAGAGATG 104  
|||||  
Qy 404 GTGAAGATTTATTT-----TAGGAGGGGAGAGAGAAATAGAAAAGAAATGACATGGT 458  
|||||  
Db 103 GGACAGATATTTCTTTGAGGGAGAGAGAGATAGAAATAGAAAAGAAATGACATGGT 44  
|||||  
Qy 459 AATCTGAAGAGATCAATTTGTTTAAAGATGAAGAGAGAAAAGA 501  
|||||  
Db 43 AGTCTCAAGATACATCGTGTTCAGACCGCAGAGAGAAAAGA 1  
|||||

## RESULT 9

## BH814587

## LOCUS

BH814587 237 bp DNA linear GSS 02-MAY-2002  
SALK\_066674 Arabidopsis thaliana TDNA insertion lines Arabidopsis  
thaliana genomic clone SALK\_066674, genomic survey sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

BH814587.1 GI:20394318  
GSS.  
Arabidopsis thaliana (thale cress)  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis  
1 (bases 1 to 237)

## REFERENCE

## AUTHORS

Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R.,  
Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L.,  
Shinn, P., Zimmerman, J. and Ecker, J.R.

## TITLE

A Sequence-Indexed Library of Insertion Mutations in the  
Arabidopsis Genome  
Unpublished (2001)

## COMMENT

Contact: Joseph R. Ecker  
Salk Institute Genomic Analysis Laboratory (SIGNAL)  
The Salk Institute for Biological Studies  
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
Tel: 858 453 4100 x1752  
Fax: 858 558 6379  
Email: ecker@salk.edu

This is single pass sequence recovered from the left border of  
TDNA.

FEATURES  
source

Class: TDNA tagged.  
Location/Qualifiers  
1..237  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/ecotype="Col-0"  
/db\_xref="taxon:3702"  
/clone="SALK\_066674"  
/clone\_lib="Arabidopsis thaliana TDNA insertion lines"  
/note="PCR was performed on Arabidopsis thaliana lines  
each of which contains one or more TDNA insertion  
elements. The resultant fragment for each line was  
directly sequenced to determine the genomic sequence at  
the site of insertion. Details of the protocols used can  
be found at [http://signal.salk.edu/tdna\\_protocols.html](http://signal.salk.edu/tdna_protocols.html)"

## ORIGIN

Query Match 11.5%; Score 110.6; DB 8; Length 237;  
Best Local Similarity 92.8%; Pred. No. 2.8e-10;  
Matches 116; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
  
Qy 833 ATATCGACCGCTAAGTTCCACCGCGGTATAAGAGTTTCTATTATAAATTTTAGCAA 892  
|||||  
Db 1 ATATCGACCGCTAAGTTCCACCGCGGTATGAGAGTTTCTATTATAAATTTTAGCAA 60  
|||||  
Qy 893 AATAAAATCAGCAAAATAATTTTTTCTGACTAAGCTTAAACGACGCCGTAAACATTTTCT 952  
|||||  
Db 61 AATAAAATCAGCAAAATAATTTTTTCTGACTAAGCTGAAAGCGGCGGTAACATTTCT 120  
|||||  
Qy 953 TCTGG 957  
|||||  
Db 121 TGTGG 125  
|||||

## RESULT 10

## BZ352865

## LOCUS

BZ352865 147 bp DNA linear GSS 14-NOV-2002  
SALK\_082722.32.45.x Arabidopsis thaliana TDNA insertion lines  
Arabidopsis thaliana genomic clone SALK\_082722.32.45.x, genomic  
survey sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

BZ352865.1 GI:24943727  
GSS.  
Arabidopsis thaliana (thale cress)  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis  
1 (bases 1 to 147)

## REFERENCE

## AUTHORS

Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R.,  
Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L.,  
Shinn, P., Zimmerman, J. and Ecker, J.R.

## TITLE

## JOURNAL

## COMMENT

A Sequence-Indexed Library of Insertion Mutations in the  
Arabidopsis Genome  
Unpublished (2001)  
Contact: Joseph R. Ecker  
Salk Institute Genomic Analysis Laboratory (SIGNAL)  
The Salk Institute for Biological Studies  
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
Tel: 858 453 4100 x1752  
Fax: 858 558 6379  
Email: ecker@salk.edu  
This is single pass sequence recovered from the left border of  
TDNA. This sequence lies within 300 bases of the 5' end of

AT3925882.  
 Class: TDNA tagged.  
 Location/Qualifiers  
 1. 147  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /ecotype="Col-0"  
 /db\_xref="taxon:3702"  
 /clone="SALK\_082722.32.45.x"  
 /notes="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at [http://signal.salk.edu/tdna\\_protocols.html](http://signal.salk.edu/tdna_protocols.html)"

ORIGIN  
 Query Match 10.5%; Score 101.2; DB 8; Length 147;  
 Best Local Similarity 92.7%; Pred. No. 1.5e-08;  
 Matches 139; Conservative 0; Mismatches 8; Indels 3; Gaps 3;  
 QY 779 ATGATGATGTCATTAGCATCAATCTCCACCGTCCCAATTTATTTAGTTGTTGACAAATATCG 838  
 Db 1 ATGATGATGTTATAGTTTGAATCTCCCTGTCCTCAATTT-TTACTTTGTTGA-AATATCG 58  
 QY 839 ACCGCTTAAGTTCACACCGACGGCTATAAGAGTTTCATTATATAAATTTTAGCAAAATATAA 898  
 Db 59 ACTGCTTAAGTTCAC-CCGACGGCTATAAGAGTTTCATTATATAAATTTTAGCAAAATATAA 117  
 QY 899 ATCAGCAAAATTAATTTTCTTGACTAAGCT 928  
 Db 118 ATCAGCAAAATTAATTTTCTTGACTAAGCT 147

RESULT 11  
 CNS0021J 1101 bp DNA linear GSS 03-JUN-1999  
 Drosophila melanogaster genome survey sequence TET3 end of BAC #  
 BACR05N11 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
 AL061936  
 AL061936.1 GI:4940214  
 GSS.  
 Drosophila melanogaster (fruit fly)  
 Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 1101)  
 Genoscope.  
 Direct Submission  
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : [secref@genoscope.cns.fr](mailto:secref@genoscope.cns.fr))  
 - Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).  
 Location/Qualifiers  
 1. 1101  
 /organism="Drosophila melanogaster"

FEATURES  
 source

FEATURES  
 source

COMMENT

/mol\_type="genomic DNA"  
 /db\_xref="taxon:7227"  
 /clone="BACR05N11"  
 /clone\_lib="RPCI-98"  
 /note="end : TET3"

ORIGIN  
 Query Match 10.4%; Score 99.8; DB 9; Length 1101;  
 Best Local Similarity 43.0%; Pred. No. 2.5e-08;  
 Matches 278; Conservative 21; Mismatches 347; Indels 0; Gaps 0;  
 QY 96 TTATCATGGCTTTTATTGATTAACATGACAAATATTATTATTTTATTTTATTTTATTTTATTTG 155  
 Db 19 KTTTNTTNTTGGTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 78  
 QY 156 GATAGCATGCAAAATATAATATATATATCAGTGTGTTAATAACATGTTTGTCTCTTAAATACAT 215  
 Db 79 NNAANANWKAATNATATTTTAAATTAATAAKAKWDTAATAAANAHAHAHAHAHA 138  
 QY 216 GCATTTTAAATCAGACATTTGTTTTTAAATCAAAATCTTAATCTCTTTATATCAACGACA 275  
 Db 139 GTGTTTTTATGAAAGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 198  
 QY 276 TTGACGGAAATTCAGTAAAGAGAAATTAAGAAATGAGAGATAGAGAGATTTCTATG 335  
 Db 199 AAAAAWAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 258  
 QY 336 GAAAAAGAAAGAGAGAACATGTAGGTGAACAAATTAAGAGATATGATGATATATTTTAT 395  
 Db 259 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 318  
 QY 396 GAGAGGTGTGAAGATTTTATTAGGAGGGAGAGAAATAGAAAAAGAAATGACATG 455  
 Db 319 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 378  
 QY 456 GTGAATCTGAAGAGATGATTCGTGTTAAAGATGAAGAGAGAAAGAGAACTCCATGGCTA 515  
 Db 379 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 438  
 QY 516 AAGTCTCGTAAAGAGATGAAAAAGAAACAAAGAGAGAGAAAGAGAGAAAGAGGCTAAA 575  
 Db 439 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 498  
 QY 576 ATAGACTAACTATGCGAAAAATTTCTGTAGCCGACAAATCTATTGTTGTTCCAAAGTTATT 635  
 Db 499 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 558  
 QY 636 TTGTGTATTTCTTTTGAAGTCAAAAGTTATTCTTACATATACTCTAAATATATAGCCGAT 695  
 Db 559 WTTAAATTTTAAATTTTAAWAAATTTTAAATAAAWATTTTAAATTTTAAATTTTAAWAAA 618  
 QY 696 ACCAATTTTCCACATCGACTCTCTCTTTATTTCCAAAAGTCAATAA 741  
 Db 619 AAAAAATTTTAAAWATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAWAAAAA 664

RESULT 12  
 BZ576813/c  
 LOCUS  
 DEFINITION  
 BZ576813  
 msh2\_5115.x1 msh Pseudomonas aeruginosa genomic clone msh2\_5115,  
 genomic survey sequence.  
 BZ576813  
 BZ576813.1 GI:27211874  
 GSS.  
 Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 1 (bases 1 to 1434)  
 Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
 Burns,J.L., Kaul,R. and Olsen,M.V.  
 Whole-Genome-Sequence variation among multiple isolates of  
 Pseudomonas aeruginosa library

REFERENCE  
 AUTHORS  
 TITLE



Db

```
Qy 394 ATGAGAGGTGGTGAAGATTATTTTATGAGAGGAGGAGAGAAATAGAAAAAGAAAAATGACA 453
Db 370 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 429
Qy 454 TGGTGAATCTGAAGAGATCAATTTGTGTTAAAGATGAAGAGAGAAAGAGAACTCCATGCC 513
Db 430 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 489
Qy 514 TAAAGTCTCTGAAGAGATGAAAGAAACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGCTA 573
Db 490 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 549
Qy 574 AAATAGACTAACTATTGCCCCAAA 596
Db 550 AAAAAAAAAAAAAAAAAACCAACAAA 572

RESULT 14
CNS0167M/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence T7 end of BAC
BACN15M24 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
VERSION AL106396.1 GI:5621701
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster (fruit fly)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1201)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sefre@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
FEATURES
source
1..1201
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN15M24"
/clone_lib="DrosBAC"
/plasmid="pBelOBAC11"
/note="end : T7"

ORIGIN
Query Match 10.1%; Score 97.2; DB 9; Length 1201;
Best Local Similarity 36.8%; Pred. No. 7.4e-08;
Matches 228; Conservative 84; Mismatches 307; Indels 0; Gaps 0;

Qy 11 AGAAATAGCAGAAATATTTTATAAAAGCATGCAATTCCTTATAGATCGCGAAGTTAA 70
Db 1155 ATANNAANANTATATAAAAWAAAAATATATATATAAAWAAAAWAAAAWATATATAATAWAA 1096
Qy 71 AAAACATATAGAAATGTTCAATATACATGCGGTTTTTATGATCAACATGACAAATAT 130
Db 1095 TATAANWAAAAAATAAAWAAAAWAAAAWTTTTHANAAATATTTTWTNATATAWAAATTTT 1036
Qy 131 TTATTTATTTTCATGAGTTTTTATTTGATGATGACAAATATTAATATATATATATATAT 190
Db 1035 TTTTATTTTATATAWAAAAWAAAAWAAAAATTTTAAAAAATAAAWTAATATATAWAAAAATTT 976
```

```
Qy 191 AACATGTTTTTGTCTTAAATATACATGCATTTTAAATATCAGACATTTGTTTTTAAATCAAA 250
Db 975 TAAAAATTTTWTATWTTTTTAAAAAATAAAATATWAAAWTTTTTTTTTATATATATAA 916
Qy 251 TCTAAATCTCTTATATACAAACGACATTGACGGAAAAATTCAGGTAAAAAGAGAAAAATAAAG 310
Db 915 WAWAWTTTTTTTTTAAAAAATAAAATTTAAAAATWTAAAAATTTATATAAATAAAAAAAA 856
Qy 311 AATGAGAGATAGAGAAATTTCTATGAAAAAAGAAAGAGAGAAACATGTAGGTGGAACAAAT 370
Db 855 AAAAAAATAAAWAAAAWAAAAATTTTAAATATATAAATTTTAAAAAATAAAATAAATAAAT 796
Qy 371 AAAGAGATATGATATATTTTATCAGAGGTGCTGAAAGATTTATTTTAGGAGGAGGAGAG 430
Db 795 TTTTAAAAATAAAAAAATAAAWAAAAAATAATTTTAAATATATAAATTTTAAATNCATAAACA 736
Qy 431 AGAAATAGAAAAAGAAAAATGCATGCTGTAATCTGAAGAAGATGAATTTGCTTTAAAGATGA 490
Db 735 AAAAAWATTTWATATAAAAAAATAAAAAAATAAAAAAATAAAAAATTTTTHACAAAMAM 676
Qy 491 AGAGAGAAAGAGAACTCCATCGCTAAAGTCTCTGTAAGAAGATGAAAAAGAAACAAAGA 550
Db 675 MAMMYMMCAAAAMAAVMAAAAMACMAMMTAAAGCNTAAAAAACAACAAAMAAAMAM 616
Qy 551 AGGAAGAGAAAGAGAAAGCGCTAAATAGACTAACTATTGCCAAAAATTTCTGTAGCGGAC 610
Db 615 AAATAAAMAAATCTAKKNDKKKAAADAKDDNKAASAMGBKKMMAATWACACTGGCGCAB 556
Qy 611 AAATACTATTTTGGTCCAAAG 629
Db 555 AMAWTTGWATTGGCAAAWS 537

RESULT 15
BH183498/c
LOCUS
DEFINITION
023_L_07-rev SmbAC1 Schistosoma mansoni genomic clone 023L07 5',
genomic survey sequence.
ACCESSION
VERSION BH183498
KEYWORDS BH183498.1 GI:16288814
SOURCE
Schistosoma mansoni
Schistosoma mansoni
Schistosoma mansoni
REFERENCE
1 (bases 1 to 660)
Le Paslier,M.C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W.,
Williams,D.L., Johnston,D., Loverde,P.T. and Le Paslier,D.
Construction and characterization of a Schistosoma mansoni
bacterial artificial chromosome library
Genomics 65 (2), 87-94 (2000)
10783255
Other GSs: 023 L_07-21
Contact: Pierce RJ
INSERM U 167
Institut Pasteur de Lille
1 rue du Professeur A. Calmette, 59019-Lille, France
Tel: (33) (0)3 20877783
Fax: (33) (0)3 20877888
Email: Raymond.Pierce@pasteur-lille.fr
CNS sequencing ID=DG0AA023CF04BP1
Plate: 023 row: L column: 07
Seq primer: M13 reverse primer
Class: BAC ends
High quality sequence stop: 660.
FEATURES
source
1..660
/organism="Schistosoma mansoni"
/mol_type="genomic DNA"
/strain="Puerto-Rican"
/db_xref="taxon:6183"
/clone="023L07"
```



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2005, 05:19:40 ; Search time 976 Seconds  
(without alignments)  
8151.195 Million cell updates/sec

Title: US-10-800-161-24  
Perfect score: 962  
Sequence: 1 tgggtttctcagaatagca.....aacattttcttctggttaac 962

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

1:	/cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2:	/cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3:	/cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4:	/cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5:	/cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6:	/cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7:	/cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8:	/cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
9:	/cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10:	/cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11:	/cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13:	/cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
14:	/cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
15:	/cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
16:	/cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
17:	/cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
18:	/cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
19:	/cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
20:	/cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq.*
21:	/cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq.*
22:	/cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq.*
23:	/cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
24:	/cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq2.*
25:	/cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*
26:	/cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
27:	/cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
28:	/cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	962	100.0	962	20	US-10-800-161-24
2	862	89.6	862	20	US-10-800-161-25
3	862	89.6	1700	20	US-10-800-161-3
4	862	89.6	1700	21	US-10-760-752-3
5	544	56.5	544	20	US-10-800-161-27
					Sequence 25, Appl
					Sequence 24, Appl
					Sequence 3, Appl
					Sequence 3, Appl
					Sequence 27, Appl

6	274	28.5	274	20	US-10-800-161-26	Sequence 26, Appl
7	272.4	28.3	274	20	US-10-800-161-28	Sequence 28, Appl
8	86	8.9	537	20	US-10-021-323-8120	Sequence 8120, Ap
9	81	8.4	1214	19	US-10-424-599-102083	Sequence 102083,
10	80.8	8.4	3673778	17	US-10-312-841-2	Sequence 2, Appli
11	79.6	8.3	2512	13	US-09-925-065A-27191	Sequence 27191, A
12	79.6	8.3	2512	13	US-09-925-065A-27192	Sequence 27192, A
13	79.2	8.2	2512	13	US-09-925-065A-27190	Sequence 27190, A
14	77.8	8.1	921	21	US-10-425-115-38710	Sequence 38710, A
15	76.8	8.0	7657	15	US-10-239-676-185	Sequence 185, App
16	76.8	8.0	7657	16	US-10-311-455-1995	Sequence 1995, Ap
17	76.8	8.0	14006	16	US-10-311-455-1931	Sequence 1931, Ap
18	76.8	8.0	1223197	14	US-10-027-632-179264	Sequence 179264,
19	76.6	8.0	1223197	18	US-10-027-632-179264	Sequence 179264,
20	76.6	8.0	529	20	US-10-021-323-10810	Sequence 10810, A
21	74.8	7.8	9539	15	US-10-239-676-52	Sequence 52, Appl
22	74.8	7.8	9539	16	US-10-240-453-54	Sequence 54, Appl
23	74.4	7.7	627	20	US-10-021-323-9336	Sequence 9336, Ap
24	74.4	7.7	2520	18	US-10-389-566-205	Sequence 205, App
25	74.2	7.7	517	20	US-10-021-323-11054	Sequence 11054, A
26	73.6	7.7	529	9	US-09-983-965-2109	Sequence 2109, Ap
27	73.6	7.7	1297	21	US-10-425-115-67687	Sequence 67687, A
28	73.4	7.6	113515	16	US-10-311-455-2147	Sequence 2147, Ap
29	73.2	7.6	478	20	US-10-021-323-6774	Sequence 6774, Ap
30	73	7.6	1029	21	US-10-425-115-123086	Sequence 123086,
31	72.8	7.6	1204	20	US-10-437-963-77858	Sequence 77858, A
32	72.6	7.5	16033	16	US-10-311-455-1377	Sequence 1377, Ap
33	72.4	7.5	803	20	US-10-437-963-72176	Sequence 72176, A
34	72.4	7.5	3673778	17	US-10-312-841-1	Sequence 1, Appli
35	72.2	7.5	778	21	US-10-363-348A-2179	Sequence 2179, Ap
36	72.2	7.5	778	21	US-10-363-348A-2180	Sequence 2180, Ap
37	72.2	7.5	778	22	US-10-363-483A-2179	Sequence 2179, Ap
38	72.2	7.5	778	22	US-10-363-483A-2180	Sequence 2180, Ap
39	72	7.5	446	9	US-09-960-352-3400	Sequence 3400, Ap
40	72	7.5	16167	16	US-10-311-455-1056	Sequence 1056, Ap
41	72	7.5	16167	16	US-10-240-485-82	Sequence 82, Appl
42	71.6	7.4	996	21	US-10-425-115-151933	Sequence 151933,
43	71.6	7.4	17934	16	US-10-311-455-1692	Sequence 1692, Ap
44	71.6	7.4	462586	24	US-10-476-264-420	Sequence 420, App
45	71.4	7.4	1759	13	US-09-925-065A-47393	Sequence 47393, A

## ALIGNMENTS

RESULT 1  
US-10-800-161-24  
; Sequence 24, Application US/10800161  
; Publication No. US20040154051A1  
; GENERAL INFORMATION:  
; APPLICANT: Cade, Rebecca M  
; APPLICANT: Dietrich, Robert A  
; APPLICANT: Lawton, Kay Ann  
; TITLE OF INVENTION: INDUCIBLE PROMOTERS  
; FILE REFERENCE: A-31089CIP1  
; CURRENT APPLICATION NUMBER: US/10/800,161  
; CURRENT FILING DATE: 2004-03-12  
; PRIOR APPLICATION NUMBER: 60/171,008  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: 60/175,519  
; PRIOR FILING DATE: 2000-01-11  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 24  
; LENGTH: 962  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-10-800-161-24

Query Match 100.0%; Score 962; DB 20; Length 962;  
Best Local Similarity 100.0%; Pred. No. 2.7e-167;  
Matches 962; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 TGTGTTCTTCAGAAATAGCAGCAAAATATTATTAAGAAGCATGCAATTCCTTTATAGATCG 60
Db 1 TGTGTTCTTCAGAAATAGCAGCAAAATATTATTAAGAAGCATGCAATTCCTTTATAGATCG 60
QY 61 CGAAGTTTAAAAAACAATAGAAATTTGTTACAATATTACATGGGTTTTTTATTGGATAACA 120
Db 61 CGAAGTTTAAAAAACAATAGAAATTTGTTACAATATTACATGGGTTTTTTATTGGATAACA 120
QY 121 TGACAAATATTATTTATTTTCATGAGTTTTTTATGGATAGCATGACAAATATTAAATATAT 180
Db 121 TGACAAATATTATTTATTTTCATGAGTTTTTTATGGATAGCATGACAAATATTAAATATAT 180
QY 181 CAGTGTTAAATACATGTTTGTCTTAAATACATGCAATTTTAAATTCAGACATTTGTTT 240
Db 181 CAGTGTTAAATACATGTTTGTCTTAAATACATGCAATTTTAAATTCAGACATTTGTTT 240
QY 241 TAAATCAAATCTAAATCTCTTATATCACACGACATTTGACGGAAAAATTCAGGTAAAAAGA 300
Db 241 TAAATCAAATCTAAATCTCTTATATCACACGACATTTGACGGAAAAATTCAGGTAAAAAGA 300
QY 301 GAAAAATAAAGATGAGAGATAGAGAGATTTCTATGAAAAAAGAAAGAGAGAAACATGTAGG 360
Db 301 GAAAAATAAAGATGAGAGATAGAGAGATTTCTATGAAAAAAGAAAGAGAGAAACATGTAGG 360
QY 361 TGAACAAATAAAGAGATATGATGATATTTTATGAGAGGTGGTGAAGATTTATTAGG 420
Db 361 TGAACAAATAAAGAGATATGATGATATTTTATGAGAGGTGGTGAAGATTTATTAGG 420
QY 421 AGAGGAGAGAGAAAAATAGAAAAAGAAAAATGACATGGTGAATCTGAAAGAGATGAAATTGTG 480
Db 421 AGAGGAGAGAGAAAAATAGAAAAAGAAAAATGACATGGTGAATCTGAAAGAGATGAAATTGTG 480
QY 481 TTAAGATGAAGAGAGAAAGAGAACTCCATGGCTAAAGTCTCGTAAAGAGATGAAAAAG 540
Db 481 TTAAGATGAAGAGAGAAAGAGAACTCCATGGCTAAAGTCTCGTAAAGAGATGAAAAAG 540
QY 541 AACAAGAGAGAGAGAAAGAGAAAGCTAAATAGACTAATCTATTTGCTCCCAAAATTTTC 600
Db 541 AACAAGAGAGAGAGAAAGAGAAAGCTAAATAGACTAATCTATTTGCTCCCAAAATTTTC 600
QY 601 TGTAGCGCAAAATACATTTTGTGTCCTCAAGGTTATTTTGTGATTTCTTTGAAAGTCAAAAG 660
Db 601 TGTAGCGCAAAATACATTTTGTGTCCTCAAGGTTATTTTGTGATTTCTTTGAAAGTCAAAAG 660
QY 661 TTATTTCTTACATATACTCTAAAAATATAGCCGATACCAATTTTCCACACATGGACTTC 720
Db 661 TTATTTCTTACATATACTCTAAAAATATAGCCGATACCAATTTTCCACACATGGACTTC 720
QY 721 CTTTATTTCCAAAAGTCAATAAAGTGTGACGTGATGATCTTACGCTTTAAAAACATCGCAT 780
Db 721 CTTTATTTCCAAAAGTCAATAAAGTGTGACGTGATGATCTTACGCTTTAAAAACATCGCAT 780
QY 781 GATGATGTCTATTAGCATCAATCTCCACCGTCCAAATTTATTTAGTTGTGCAATATCGAC 840
Db 781 GATGATGTCTATTAGCATCAATCTCCACCGTCCAAATTTATTTAGTTGTGCAATATCGAC 840
QY 841 CGTCTAAGTTCCACACGCGGTATAAGAGTTTCATATAAAATTTTAGCAAAATAAAAT 900
Db 841 CGTCTAAGTTCCACACGCGGTATAAGAGTTTCATATAAAATTTTAGCAAAATAAAAT 900
QY 901 CAGCAATAATTTTCTTCACTAAGCTTAAACGAGCCGTTAAACATTTTCTTCTGGCTA 960
Db 901 CAGCAATAATTTTCTTCACTAAGCTTAAACGAGCCGTTAAACATTTTCTTCTGGCTA 960
QY 961 AC 962
Db 961 AC 962
```

RESULT 2  
US-10-800-161-25  
; Sequence 25, Application US/10800161  
; Publication No. US20040154051A1

```
; GENERAL INFORMATION:  
; APPLICANT: Cade, Rebecca M  
; APPLICANT: Dietrich, Robert A  
; APPLICANT: Lawton, Kay Ann  
; TITLE OF INVENTION: INDUCIBLE PROMOTERS  
; FILE REFERENCE: A-31089CIP1  
; CURRENT APPLICATION NUMBER: US/10/800,161  
; CURRENT FILING DATE: 2004-03-12  
; PRIOR APPLICATION NUMBER: 60/171,008  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: 60/175,519  
; PRIOR FILING DATE: 2000-01-11  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 25  
; LENGTH: 862  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; US-10-800-161-25
```

Query Match 89.6%; Score 862; DB 20; Length 862;  
Best Local Similarity 100.0%; Pred. No. 6.5e-149;  
Matches 862; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 101 TGGGTTTTTATTCGGATAACATGACAAATATTTTATTTTCATGAGCTTTTATTGGATAG 160
Db 1 TGGGTTTTTATTCGGATAACATGACAAATATTTTATTTTCATGAGCTTTTATTGGATAG 60
QY 161 CATGACAAATATTAAATATATATCATGTTGTTAAATCAATCTTTTGTTCCTTAAAAATACATGCATT 220
Db 61 CATGACAAATATTAAATATATATCATGTTGTTAAATCAATCTTTTGTTCCTTAAAAATACATGCATT 120
QY 221 TTAATAATCAGACATTTGTTTTTAAAAATCAAAATCTAAATCTCTTATATATCAACGACATTGAC 280
Db 121 TTAATAATCAGACATTTGTTTTTAAAAATCAAAATCTAAATCTCTTATATATCAACGACATTGAC 180
QY 281 GGAATAATTCAGGTAAAGAGAAAAATAAAGATGAGAGATAGAGATTTCTATGAAAA 340
Db 181 GGAATAATTCAGGTAAAGAGAAAAATAAAGATGAGAGATAGAGATTTCTATGAAAA 240
QY 341 AGAAAGAGAGAACATCTAGGTGAAACAAAAATAAAGAGATATGATGATATATTTTATGAGAG 400
Db 241 AGAAAGAGAGAACATCTAGGTGAAACAAAAATAAAGAGATATGATGATATATTTTATGAGAG 300
QY 401 GTGGTCAAGATTTATTTTAGGAGGAGAGAGAAAAATAGAAAAAGAAAAATGACATGGTGAA 460
Db 301 GTGGTCAAGATTTATTTTAGGAGGAGAGAGAAAAATAGAAAAAGAAAAATGACATGGTGAA 360
QY 461 TCTGAAGAGAGATGAATTTGTTTAAAGATGAAGAGAGAGAACTCCATGGCTAAAGTC 520
Db 361 TCTGAAGAGAGATGAATTTGTTTAAAGATGAAGAGAGAGAACTCCATGGCTAAAGTC 420
QY 521 TCGTAAAGAGATGAAAAAGAAAAAAGAAAGAGAAAGAAAGAAAGCTAAAAATAGA 580
Db 421 TCGTAAAGAGATGAAAAAGAAAAAAGAAAGAGAAAGAAAGAAAGCTAAAAATAGA 480
QY 581 CTAACTATTTCGCAAAATTTCTGTAGCCGACAAATATCTATTTGGTCCAAGGTTATTTTGTG 640
Db 481 CTAACTATTTCGCAAAATTTCTGTAGCCGACAAATATCTATTTGGTCCAAGGTTATTTTGTG 540
QY 641 TATTCTTTTGAAGTCAAAAGTTATTTCTTACATATACTCTTAAAAATATAGCCGATACCAA 700
Db 541 TATTCTTTTGAAGTCAAAAGTTATTTCTTACATATACTCTTAAAAATATAGCCGATACCAA 600
QY 701 TTTTCCACACATGGACTTCCTTTTATTTCCAAAAGTCAATAAAGTGTGACGTCATGATPACT 760
Db 601 TTTTCCACACATGGACTTCCTTTTATTTCCAAAAGTCAATAAAGTGTGACGTCATGATPACT 660
QY 761 TAGCGTTTAAAAACATCGCATGATGTCATTTAGCATCAATCTCCACCGTCCAAATTTATT 820
Db 661 TAGCGTTTAAAAACATCGCATGATGTCATTTAGCATCAATCTCCACCGTCCAAATTTATT 720
QY 821 TAGTGTGTGACAATATCGACCGTCTAAGTTCCACACCGCGCTATTAAGAGTTTTCATTAT 880
```

[illegible]

US-10-760-752-3

```

; Sequence 3, Application US/10760752
; Publication No. US20040248303A1
; GENERAL INFORMATION:
; APPLICANT: Cade, Rebecca M
; APPLICANT: Dietrich, Robert A
; TITLE OF INVENTION: Inducible Promoter Fragment
; FILE REFERENCE: A-31089D1V
; CURRENT APPLICATION NUMBER: US/10760752
; CURRENT FILING DATE: 2004-01-20
; PRIOR APPLICATION NUMBER: 09/733685
; PRIOR FILING DATE: 2000-12-8
; PRIOR APPLICATION NUMBER: 60/171,008
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/175,519
; PRIOR FILING DATE: 2000-01-11

```

```

Query Match      89.6%; Score 862; DB 20; Length 1700;
Best Local Similarity 100.0%; Pmid.No. 8.2e-149;
Matches 862; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY      101  TGGGTTTTTATTCGATAACATGACAAATATTTATTTATTTTCATGAGTTTTTATTCGATAG 160
          |||||
Db       1   TGGGTTTTTATTCGATAACATGACAAATATTTATTTATTTTCATGAGTTTTTATTCGATAG 60
QY      161  CATGACAAATATTAAATATATCAGTGTGTAATAACATGTTTTTGTTCTTAAATAACATGCATT 220

```





Db 241 TTGAAGTCAAAAGTATTCTTACATATACCTCTAAAAATATAGCCGATACCAATTTTTC 300  
Qy 708 ACATCGACTTCCTTTATTCCTAAAGTCAATTAAGTGTGACGTCATGATACCTTACGCTT 767  
Db 301 ACATCGACTTCCTTTATTCCTAAAGTCAATTAAGTGTGACGTCATGATACCTTACGCTT 360  
Qy 768 TAAACATCGCATGATGATGTCATTAGCATCAATCTCCACCGTCCAAATTTTATTTAGTTGT 827  
Db 361 TAAACATCGCATGATGATGTCATTAGCATCAATCTCCACCGTCCAAATTTTATTTAGTTGT 420  
Qy 828 TGCAATATCGACCGTCTAAGTTCCACACGACCGCTTAAAGATTTTCAATATAAATTTT 887  
Db 421 TGCAATATCGACCGTCTAAGTTCCACACGACCGCTTAAAGATTTTCAATATAAATTTT 480  
Qy 888 AGCAAAATAAATCAGCAAAATAATTTTCTTGACTAAGCTTAAACGACCGGTTAACAT 947  
Db 481 AGCAAAATAAATCAGCAAAATAATTTTCTTGACTAAGCTTAAACGACCGGTTAACAT 540  
Qy 948 TTTC 951  
Db 541 TTTC 544

RESULT 6  
US-10-800-161-26  
; Sequence 26, Application US/10800161  
; Publication No. US20040154051A1  
; GENERAL INFORMATION:  
; APPLICANT: Cade, Rebecca M  
; APPLICANT: Dietrich, Robert A  
; APPLICANT: Lawton, Kay Ann  
; TITLE OF INVENTION: INDUCIBLE PROMOTERS  
; FILE REFERENCE: A-31089CIP1  
; CURRENT APPLICATION NUMBER: US/10/800,161  
; CURRENT FILING DATE: 2004-03-12  
; PRIOR APPLICATION NUMBER: 60/171,008  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: 60/175,519  
; PRIOR FILING DATE: 2000-01-11  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 26  
; LENGTH: 274  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-10-800-161-26

Query Match 28.5%; Score 274; DB 20; Length 274;  
Best Local Similarity 100.0%; Pred. No. 7.5e-41;  
Matches 274; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 678 TCTAAAAATATAGCCGATACCAATTTTCCACACATGCGCTTCTTTATTCCTCAAAAGTCA 737  
Db 1 TCTAAAAATATAGCCGATACCAATTTTCCACACATGCGCTTCTTTATTCCTCAAAAGTCA 60  
Qy 738 ATAAAGTGTGACGTCATGATGATGTCATTAGCTTTTAAACATCGCATGATGCTCATTAGCAT 797  
Db 61 ATAAAGTGTGACGTCATGATGATGTCATTAGCTTTTAAACATCGCATGATGCTCATTAGCAT 120  
Qy 798 CAATCTCCACCGTCCCAATTTATTTAGTTGTTGACAATATCGACCGTCTAAAGTTCCACACC 857  
Db 121 CAATCTCCACCGTCCCAATTTATTTAGTTGTTGACAATATCGACCGTCTAAAGTTCCACACC 180  
Qy 858 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAATCAGCAAAATAAATTTTTC 917  
Db 181 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAATCAGCAAAATAAATTTTTC 240  
Qy 918 TTGACTAAGCTTAAACGACCGGTTAACATTTTC 951  
Db 241 TTGACTAAGCTTAAACGACCGGTTAACATTTTC 274

RESULT 7  
US-10-800-161-28  
; Sequence 28, Application US/10800161  
; Publication No. US20040154051A1  
; GENERAL INFORMATION:  
; APPLICANT: Cade, Rebecca M  
; APPLICANT: Dietrich, Robert A  
; APPLICANT: Lawton, Kay Ann  
; TITLE OF INVENTION: INDUCIBLE PROMOTERS  
; FILE REFERENCE: A-31089CIP1  
; CURRENT APPLICATION NUMBER: US/10/800,161  
; CURRENT FILING DATE: 2004-03-12  
; PRIOR APPLICATION NUMBER: 60/171,008  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: 60/175,519  
; PRIOR FILING DATE: 2000-01-11  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 28  
; LENGTH: 274  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-10-800-161-28

Query Match 28.3%; Score 272.4; DB 20; Length 274;  
Best Local Similarity 99.6%; Pred. No. 1.5e-40;  
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 678 TCTAAAAATATAGCCGATACCAATTTTCCACACATGCGCTTCTTTATTCCTCAAAAGTCA 737  
Db 1 TCTAGAAATATAGCCGATACCAATTTTCCACACATGCGCTTCTTTATTCCTCAAAAGTCA 60  
Qy 738 ATAAAGTGTGACGTCATGATGATGTCATTAGCTTTTAAACATCGCATGATGCTCATTAGCAT 797  
Db 61 ATAAAGTGTGACGTCATGATGATGTCATTAGCTTTTAAACATCGCATGATGCTCATTAGCAT 120  
Qy 798 CAATCTCCACCGTCCCAATTTATTTAGTTGTTGACAATATCGACCGTCTAAAGTTCCACACC 857  
Db 121 CAATCTCCACCGTCCCAATTTATTTAGTTGTTGACAATATCGACCGTCTAAAGTTCCACACC 180  
Qy 858 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAATCAGCAAAATAAATTTTTC 917  
Db 181 GACGGCTATAAGAGTTTCATTATAAATTTTAGCAAAATAAATCAGCAAAATAAATTTTTC 240  
Qy 918 TTGACTAAGCTTAAACGACCGGTTAACATTTTC 951  
Db 241 TTGACTAAGCTTAAACGACCGGTTAACATTTTC 274

RESULT 8  
US-10-021-323-8120  
; Sequence 8120, Application US/10021323  
; Publication No. US20040123340A1  
; GENERAL INFORMATION:  
; APPLICANT: Deikman, Jill  
; APPLICANT: Feng, Paul C. C.  
; APPLICANT: Fincher, Karen L.  
; APPLICANT: Ziegler, Todd E.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(52274)B  
; CURRENT APPLICATION NUMBER: US/10/021,323  
; CURRENT FILING DATE: 2001-12-12  
; PRIOR APPLICATION NUMBER: US 60/255, 619  
; PRIOR FILING DATE: 2000-12-14  
; NUMBER OF SEQ ID NOS: 17880  
; SEQ ID NO 8120  
; LENGTH: 537  
; TYPE: DNA  
; ORGANISM: Gossypium hirsutum  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3828-006-Q1-N6-D9  
US-10-021-323-8120





```
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27190
; LENGTH: 2512
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-27190

Query Match      8.2%; Score 79.2; DB 13; Length 2512;
Best Local Similarity 54.5%; Pred. No. 0.00011;
Matches 156; Conservative 1; Mismatches 129; Indels 0; Gaps 0;

QY 283 AAAATTTCAGGTAAAGAGAGAAATAAAGAAATGAGAGATAGAGAGATTCTTATCGAAAAAG 342
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 690 AAAAGAGAAAGAGAAAGAAAGAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAA 749
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 343 AAAGAGAGAACATGTAGGTGAAACAAAATAAAGAGATATGATATATATTTATGAGAGGT 402
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 750 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 809
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 403 GGTGAAGATTATTTAGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATC 462
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 810 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 869
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 463 TGAAGAAGATGAATTGTGTTAAAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 522
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 870 AAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 929
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 523 GTAAGAAGATGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 568
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 930 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 975

RESULT 14
US-10-425-115-38710/c
; Sequence 38710, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 38710
; LENGTH: 921
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(921)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_135307C.1
US-10-425-115-38710

Query Match      8.1%; Score 77.8; DB 21; Length 921;
Best Local Similarity 45.7%; Pred. No. 0.00014;
Matches 271; Conservative 0; Mismatches 322; Indels 0; Gaps 0;

; PRIOR APPLICATION NUMBER: US 60/239,676
; PRIOR FILING DATE: 2002-09-24
; CURRENT APPLICATION NUMBER: US/10/239,676
; FILE REFERENCE: 5013.1003
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 185
; LENGTH: 7657
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-185

QY 23 AAATATTTTAAAAAGCATGCAATTCCTTTATAGATCGCGAAGCTTTAAAAAAACATATAG 82
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 861 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAA 802
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 83 AATTGTTTACAATATTACATGGGTTTTTATTGGATACATGACAAATATTATTATTATTCA 142
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 801 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 742
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 143 TGAGTTTTTATGGATAGCATGACAATATTATATATATCAGTGTATTAACATGTTTGT 202
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 741 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 682
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 203 TCTTAAAAATACATGCATTTTAAATCAGACATTTGTTTAAATCAAAATCTAATCTCTTA 262
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 681 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 622
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 263 TATCACAACGACATTCACGGAAAAATTCAGGTAAAAAGAGAGAGAGAGAGAGAGAGATAG 322
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 621 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 562
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 323 AGAGATTTCTATGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATATGA 382
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 561 CAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 502
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 383 TGATATATTTTATGAGAGGTGTTGAAGATTTATTTTAGGAGAGGAGAGAGAGAGAGAGAA 442
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 501 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 442
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 443 AGAAATGACATGCTGAATCTGAGAGAGATGATTTGTTTAAGATGAAGAGAGAGAGAGAG 502
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 441 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 382
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 503 AACTCATGCTAAAGTCTCGTAAAGAGATGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAA 562
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 381 ACATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 322
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 563 GAGAAAGGCTAAATAAGACTAACTATTGCCAAAAATTTCTGTAGCCGACAAATA 615
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 321 CAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 269
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
US-10-239-676-185/c
; Sequence 185, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 185
; LENGTH: 7657
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-185
```



**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2005, 02:50:19 ; Search time 217 Seconds  
(without alignments)  
7253.909 Million cell updates/sec

Title: US-10-800-161-24  
Perfect score: 962  
Sequence: 1 tgtgtttctcagaatagca.....aacattttcttctggttaac 962

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	862	89.6	1700	US-09-733-685-3	Sequence 3, Appli
2	103.2	10.7	7218	US-08-232-463-14	Sequence 14, Appl
3	76	7.9	612	US-09-902-540-1357	Sequence 1357, Ap
4	76	7.9	19124	US-08-487-8268-13	Sequence 13, Appl
5	72.2	7.5	12703	US-09-949-016-16685	Sequence 16685, A
6	71.2	7.4	152331	US-09-128-155-16	Sequence 15, Appl
7	71.2	7.4	176373	US-09-128-155-17	Sequence 17, Appl
8	69	7.2	133559	US-09-949-016-15845	Sequence 15845, A
9	67.2	7.0	36731	US-09-949-016-13770	Sequence 13770, A
10	67.2	7.0	168174	US-10-071-411A-63	Sequence 63, Appl
11	67.2	7.0	168273	US-10-071-411A-2	Sequence 2, Appli
12	66	6.9	96922	US-09-949-016-17061	Sequence 17061, A
13	66	6.9	258775	US-09-949-016-16435	Sequence 16435, A
14	65.8	6.8	147382	US-09-949-016-14624	Sequence 14624, A
15	64.8	6.7	55841	US-09-949-016-16602	Sequence 16602, A
16	64.8	6.7	248968	US-09-949-016-12614	Sequence 12614, A
17	64.8	6.7	250958	US-09-949-016-16061	Sequence 16061, A
18	64.4	6.7	601	US-09-949-016-30531	Sequence 30531, A
19	64.4	6.7	601	US-09-949-016-37150	Sequence 37150, A
20	64.4	6.7	601	US-09-949-016-37150	Sequence 37150, A
21	64.4	6.7	601	US-09-949-016-145868	Sequence 145868, A
22	64.4	6.7	601	US-09-949-016-146136	Sequence 146136, A
23	64.4	6.7	601	US-09-949-016-146404	Sequence 146404, A
24	64.4	6.7	205044	US-09-949-016-15851	Sequence 15851, A
25	64.4	6.7	205044	US-09-949-016-15852	Sequence 15852, A
26	64.4	6.7	205044	US-09-949-016-15853	Sequence 15853, A
27	64.4	6.7	223471	US-09-949-016-12387	Sequence 12387, A

28	64.4	6.7	223471	4	US-09-949-016-12724	Sequence 12724, A
29	64.4	6.7	223471	4	US-09-949-016-12725	Sequence 12725, A
30	63.8	6.6	509	4	US-09-733-685-1	Sequence 1, Appli
C 31	62.6	6.5	601	4	US-09-949-016-184864	Sequence 184864, A
C 32	62.6	6.5	601	4	US-09-949-016-184865	Sequence 184865, A
C 33	62.6	6.5	601	4	US-09-949-016-184866	Sequence 184866, A
C 34	62.6	6.5	601	4	US-09-949-016-184867	Sequence 184867, A
C 35	62.6	6.5	601	4	US-09-949-016-184868	Sequence 184868, A
C 36	62.6	6.5	601	4	US-09-949-016-184869	Sequence 184869, A
C 37	62.6	6.5	601	4	US-09-949-016-184870	Sequence 184870, A
C 38	62.6	6.5	601	4	US-09-949-016-184871	Sequence 184871, A
C 39	62.6	6.5	601	4	US-09-949-016-185019	Sequence 185019, A
C 40	62.6	6.5	601	4	US-09-949-016-185020	Sequence 185020, A
C 41	62.6	6.5	601	4	US-09-949-016-185021	Sequence 185021, A
C 42	62.6	6.5	601	4	US-09-949-016-185022	Sequence 185022, A
C 43	62.6	6.5	601	4	US-09-949-016-185023	Sequence 185023, A
C 44	62.6	6.5	601	4	US-09-949-016-185024	Sequence 185024, A
C 45	62.6	6.5	601	4	US-09-949-016-185025	Sequence 185025, A

ALIGNMENTS

RESULT 1

US-09-733-685-3  
; Sequence 3, Application US/097333685  
; Patent No. 6706952  
; GENERAL INFORMATION:  
; APPLICANT: Cade, Rebecca M  
; APPLICANT: Dietrich, Robert A  
; TITLE OF INVENTION: GENES ENCODING PROTEINS INVOLVED IN THE REGULATION OF  
; FILE REFERENCE: A-31089A  
; CURRENT APPLICATION NUMBER: US/09/733,685  
; CURRENT FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/171,008  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: 60/175,519  
; PRIOR FILING DATE: 2000-01-11  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 1700  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (365)..(374)  
; OTHER INFORMATION: TCAL motif  
; NAME/KEY: misc feature  
; LOCATION: (426)..(435)  
; OTHER INFORMATION: TCAL motif  
; NAME/KEY: misc feature  
; LOCATION: (609)..(614)  
; OTHER INFORMATION: MYCATR22 element  
; NAME/KEY: misc feature  
; LOCATION: (646)..(665)  
; OTHER INFORMATION: CAMV ASI salicylic acid response element  
; NAME/KEY: misc feature  
; LOCATION: (707)..(712)  
; OTHER INFORMATION: PAL BOX  
; NAME/KEY: misc feature  
; LOCATION: (757)..(762)  
; OTHER INFORMATION: HEXAMERAT 4 element  
; NAME/KEY: misc feature  
; LOCATION: (863)..(1228)  
; OTHER INFORMATION: Nil6 genomic coding region  
; US-09-733-685-3

Query Match 89.6%; Score 862; DB 4; Length 1700;  
Best Local Similarity 100.0%; Pred. No. 5e-175;  
Matches 862; Conservative 0; Mismatches 0; Indels 0; Gaps 0;











Db 38766 ATATGTTATTTCTATTAATAATATATATGTTATTTATCTATTAATAATATATATATTAATATAT 38707  
Qy 164 GACAAATATTAAATATATACAGTGTGTTAAATAACATGTTTGTCTTAAATATACATGATTTTAA 223  
Db 38706 -AAATATATATATAAATATATAAATAATATATTTATATATTAATATATAAATATATTTT 38648  
Qy 224 AAATCAGACATTTGTTTAAATCAATCTAATCTCTTATATCAACAG-ACATGTGCGG 282  
Db 38647 ATATATTAAATATATAAATATATTTATATATCAAAATATATATAAATATATATTTATAT 38588  
Qy 283 AAAATTCAGGTAAAGAGAAATAAAGAAATGAGAGATGAGAGATTTCTATGGAAGAAAG 342  
Db 38587 ATCAAAATATATAAATATATTTATATATCAAAATATATAAATATATAAATATATCAAA 38528  
Qy 343 AAAGAGAGACATGTAGGTGAACAAAATAAAGAGATATGATGATATTTTATGAGAGGT 402  
Db 38527 TATATAAATATATTTATATATCAATATATATAAATATATTTATATATCAAAATATATA 38468  
Qy 403 GGTGAAGATTTTATGAGAGGGAGAGAGAAATAGAAATAAAGAAATGACATGCTGAATC 462  
Db 38467 TAAAAACATATTTATATATCAATATATATAAATATATTTATATATCAAAATATATA 38408  
Qy 463 TGAAGAAGATGAATGTTTAAAGATGAAGAGAGAAAGAACTCCATGCTTAAAGTCTC 522  
Db 38407 AAATATATTTATATATTTAAATATATATATAAATATATATAAATATATATAAATGATG 38348  
Qy 523 GTAAAGAGATGAAAAAGAAACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 580  
Db 38347 GCAAAACAGATAAATAAATCTGAAAGGCAAAACAGATAAATCCAAATAAACAGTGA 38290

## RESULT 13

US-09-949-016-16435  
; Sequence 16435, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 16435  
; LENGTH: 258775  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-16435

Query Match 6.9%; Score 66; DB 4; Length 258775;  
Best Local Similarity 51.7%; Pred. No. 0.00016;  
Matches 150; Conservative 0; Mismatches 140; Indels 0; Gaps 0;  
Qy 295 AAAAGAGAAATAAAGATGAGAGATGAGAGATTTCTATGGAAGAAAGAAAGAGAGAA 354  
Db 122511 AAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 122570  
Qy 355 TGTAGGTGACAAATTAAGAGATATGATGATATTTTATGAGAGGTGCTGAAGATTAT 414  
Db 122571 AGGAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 122630  
Qy 415 TTTAGGAGCGGAGAGAGAAATAGAAAGAAAGAAATGACATGCTCAAGAGATGA 474  
Db 122631 GAGAGAGAAAGAAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 122690  
Qy 475 ATTGTGTTAAAGATGAAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 534

Db 122691 AGAGGAG 122750  
Qy 535 AAAAGAGAAACAAAAG 584  
Db 122751 AGAAG 122800  
RESULT 14  
US-09-949-016-14624/c  
; Sequence 14624, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 14624  
; LENGTH: 147382  
; TYPE: DNA  
; ORGANISM: Human  
; NAME/KEY: misc feature  
; LOCATION: (1)-(147382)  
; OTHER INFORMATION: n = A, T, C or G  
US-09-949-016-14624

Query Match 6.8%; Score 65.8; DB 4; Length 147382;  
Best Local Similarity 47.1%; Pred. No. 0.00016;  
Matches 202; Conservative 0; Mismatches 227; Indels 0; Gaps 0;  
Qy 11 AGAAATAGCAGCAAAATATTTTATAAAGCATGCAATTTCTTTATAGATCCGGAAGTTTAA 70  
Db 137087 AAATATATATAAATATATATATATTAATATACATAAATATGTATAAATATAAATATAT 137028  
Qy 71 AAAACATATAGAAITGTTTACATATATACATGGTTTTTTATTCGATAACATGACAAATAT 130  
Db 137027 TAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATAT 136968  
Qy 131 TTATTTATTTTCATGAGTTTTTTATTTGGATAGCATGACAAATATTAATATATATCATGTTTAA 190  
Db 136967 ATATAAATATGTTTAATATATACATATATAAATAAATATATAAATAAATATATAAATAT 136908  
Qy 191 AACATGTTTTGTTCTTTAAATATACATGCAATTTTAAAAATCAGACATTTGTTTTAAAAATCAAA 250  
Db 136907 ATATAACTAATAATATAAATATATAAATAAATATAAATAAATATAAATAAATATATATGAC 136848  
Qy 251 TCTAATCTTATATACACACGACATGACGGAATTCAGGTAAAAAGAGAGAAATTAAG 310  
Db 136847 TAATATATAAATATATAAATATATAAATAAATATAAATAAATATAAATAAATATATACAAAT 136788  
Qy 311 AATGAGAGATAGAGAGATTTCTATGGAAGAAAGAGAGAGAAATGATGAGTGAACAAAT 370  
Db 136787 ATATATAAATATAAATATAAATAAATATAAATAAATATAAATAAATATAAATATATATAAT 136728  
Qy 371 AAAGAGATATGATATATATTTTATGAGAGGTGGTGAAGATTTATTTTAGGAGAGGAGAG 430  
Db 136727 AAATATATAAATAAATATATAAATAAATATAAATAAATATAAATAAATATAAATAAATAT 136668  
Qy 431 AGAATAGA 439  
Db 136667 ATATATATA 136659



## RESULT 15

US-09-949-016-16602/c  
; Sequence 16602, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16602  
; LENGTH: 55841  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-16602

Query Match 6.7%; Score 64.8; DB 4; Length 55841;  
Best Local Similarity 53.8%; Pred. No. 0.00021;  
Matches 155; Conservative 0; Mismatches 132; Indels 1; Gaps 1;  
  
QY 294 AAAAAGAGAAATAAAGATGAGAGATAGAGATTTCTATGGAAAAAGAAAGAGAGAAC 353  
DB 31510 AAAGAGAGAGAGAAAGGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 31451  
  
QY 354 ATGTAGGTGAACAAAAATAAGAGATATGATGATATATTTTATGAGAGGTGGTGAAGATTA 413  
DB 31450 AAAAGAAAGAAAGAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 31391  
  
QY 414 TTTTAGGAGAGGAGAGAGAAATAGAAAAAGAAAAATGACATGTTGTAATCTTGAAGAGATG 473  
DB 31390 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 31331  
  
QY 474 AATTGTTTAAAGATGAGAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAT 533  
DB 31330 GAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 31271  
  
QY 534 GAAAAAGAAACAAAAAGAGGAG-AGAAAGAGAGAAAGGCTAAAAATAGA 580  
DB 31270 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 31223

Search completed: November 7, 2005, 06:51:43  
Job time : 219 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2005, 06:51:50 ; Search time 4632 Seconds  
(without alignments)  
10063.454 Million cell updates/sec

Title: US-10-800-161-24  
Perfect score: 962  
Sequence: 1 tgggtttctcagaatagca.....aacattttcttctggttaac 962

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	962	100.0	83650	8	AB023041 Arabidops
C 2	862	89.6	1700	6	AR488147 Sequence
C 3	676.2	70.3	92620	8	AB026636 Arabidops
C 4	632.2	65.7	83646	8	AB005248 Arabidops
C 5	593.8	61.7	94487	8	AC012394 Arabidops
C 6	593.8	61.7	100806	8	AC015450 Arabidops
C 7	579.2	60.2	104386	8	AT132A17 Arabidops
C 8	579.2	60.2	179771	8	ATCHRIV25 Arabidops
C 9	571.6	59.4	95519	8	AF071527 Arabidops
C 10	571.6	59.4	116448	8	AC005142 Arabidops
C 11	571.6	59.4	159629	8	ATCHRIV9 Arabidops
C 12	406.8	42.3	95190	8	AC007203 Arabidops
C 13	120.6	12.5	105223	8	AC007399 Arabidops
C 14	103.2	10.7	7218	6	I66494 Sequence 14
C 15	99.2	10.3	19577	10	BX255966 Mouse DNA
C 16	94.2	9.8	1197	8	AJ840663 Arabidops
C 17	92.6	9.6	175366	10	AC101718 Mus muscu
C 18	90.2	9.4	173049	10	AC115878 Mus muscu
C 19	90	9.4	286208	2	AC117140 Rattus no

C 20	89.2	9.3	191209	10	AC125184
C 21	88.8	9.2	171258	2	AC116819
C 22	87.6	9.1	164443	2	CR751221
C 23	87.4	9.1	161115	2	AC101939
C 24	87.4	9.1	174536	10	AC125079
C 25	87.4	9.1	216464	10	AL844515
C 26	86.8	9.0	191001	2	AC138586
C 27	86.8	9.0	202872	2	AC016160
C 28	86.6	9.0	165720	10	AC127244
C 29	86.4	9.0	110000	8	CR382131.03
C 30	86.2	9.0	218313	10	AC132465
C 31	86	8.9	175787	10	AC142103
C 32	85.6	8.9	43736	9	AC130469
C 33	85.6	8.9	44735	9	AC092315
C 34	85.6	8.9	64789	2	AC083839
C 35	85.6	8.9	227625	2	AC094806
C 36	85.6	8.9	239189	2	AC097180
C 37	85	8.8	178038	10	AL672308
C 38	84.6	8.8	422	5	PDOMU4
C 39	84.6	8.8	164094	2	AC101966
C 40	84.6	8.8	176666	2	AC132131
C 41	84.6	8.8	177649	10	AL732328
C 42	84.4	8.8	159453	10	AC126558
C 43	84.4	8.8	224248	2	AC124316
C 44	84.2	8.8	245524	2	AC095456
C 45	84.2	8.8	270712	2	AC137360

## ALIGNMENTS

RESULT 1	AB023041/c
LOCUS	AB023041 83650 bp DNA linear PLN 14-FEB-2004
DEFINITION	Arabidopsis thaliana genomic DNA, chromosome 3, PI clone: MPE11.
ACCESSION	AB023041 BA000014
VERSION	AB023041.1 GI:4220640
KEYWORDS	Arabidopsis thaliana (thale cress)
SOURCE	Arabidopsis thaliana
ORGANISM	Arabidopsis thaliana
REFERENCE	1 Sato, S., Nakamura, Y., Kaneko, T., Katoh, T., Asamizu, E. and Tabata, S. Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence features of the regions of 4,504,864 bp covered by sixty PI and TAC clones
JOURNAL	DNA Res. 7 (2), 131-135 (2000)
MEDLINE	20277480
PUBMED	10819329
REFERENCE	2 (bases 1 to 83650)
AUTHORS	Sato, S., Nakamura, Y., Kaneko, T., Katoh, T., Asamizu, E. and Tabata, S.
TITLE	Direct Submission
JOURNAL	Submitted (01-FEB-1999) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp, Tel: 81-438-52-3935, Fax: 81-438-52-3934)
COMMENT	Address for correspondence: kaos@kazusa.or.jp For the latest information on annotation of this clone, please see <a href="http://www.kazusa.or.jp/kaos/cgi-bin/sgd/graph.cgi?c=MPE11">http://www.kazusa.or.jp/kaos/cgi-bin/sgd/graph.cgi?c=MPE11</a> Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, <a href="http://compbio.ornl.gov/Grail-1.3/">http://compbio.ornl.gov/Grail-1.3/</a> ), GENSCAN (Chris Burge, MIT, <a href="http://CCR-081.mit.edu/GENSCAN.html">http://CCR-081.mit.edu/GENSCAN.html</a> ), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <a href="http://www.cbs.dtu.dk/services/NetGene2/">http://www.cbs.dtu.dk/services/NetGene2/</a> ) and SplicePredictor (Volker Brendel, Stanford University, <a href="http://gremlin.zoel.iastate.edu/cgi-bin/sp.cgi">http://gremlin.zoel.iastate.edu/cgi-bin/sp.cgi</a> ). Genes encoding tRNAs are predicted by tRNAscan-SE

(Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAcan-SE/). of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is K9122 and the 3' clone is MUL14.

FEATURES

Source

1. .83650  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:3702"  
/chromosome="3"  
/clone="MPE11"  
/clone\_lib="Miteui P1"  
/ecotype="Columbia"  
/notes="complement(342. .764)  
/notes="CDS is reported in Acc# AP000599  
contains similarity to CHP-rich zinc finger protein  
gene id:K9122.5"  
/number=1  
/evidence=not\_experimental  
join(1705. .2463,2548. .2739,2862. .2949,3037. .3107,  
3190. .3273,3410. .3514)  
/notes="gb|AAD55139.1  
gene id:MPE11.1"  
/codon\_start=1  
/evidence=not\_experimental  
/product="dihydroliipoamide S-acetyltransferase"  
/protein\_id="BAB01047.1"  
/db\_xref="GI:9279589"  
/translation="MTVRSKIRIFMPALSSSTMTEGKIVSWIKTEGEKLAKGESVVVV  
ESDKADMETFDVGGVLAIVVEGETAPVGAIGLAEAEIEAEKKAASKSSSS  
VBAVVPSPPTVSSAPALAPAPVATVSDGPRKVTAPYAKLAKQKHQVIESVAG  
TGFGRTASDVTAAGIAPSKSIAPPPPPPTAKATNTNLPPLPDSIVPPTA  
MOSVAKNMIESLVPFRVGVVNTDALDYKVKPKGVTMTALLAKAAGMALAQH  
PVNASKDKGSFYSNINIAVAINGLIIPVLQADKLDLYLLSOKWELVGKA  
RSKLOPHYNSGTFTLSNLGMFQVDRFDAILPPGOGAIMAVGASKPTTVADKGFSS  
KQNTMLVNTADHRIYVGADLAFLQTFAKLIENPDSLIL"  
complement(4594. .5106)  
/notes="unnamed protein product; gene\_id:MPE11.2  
unknown protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB01048.1"  
/db\_xref="GI:9279590"  
/translation="MEDLEEERLTRDTSVGNKRVRGCLDLDSPVVKRLRDDILFDSSGL  
DPVSQDLDSVKMSFENELSTTTAALSGETQPDGLYLFASDDELGLPPLTPQTLL  
PSCSETVTELVRASDSSEVGELCGFEDHVTFFGCDLGDGLFFYFDGCLDSGDLF  
SWRPEFLPAE"  
complement(join(10090. .10408,10649. .10804,10916. .10962))  
/notes="unnamed protein product; gene\_id:MPE11.3  
similar to unknown protein  
sp|P42744  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB01049.1"  
/db\_xref="GI:9279591"  
/translation="MMPEKAKYDRQLMYTIQGTLEBASICLLNCGPIGNSALKNLVLG  
GVGSIITVSGSKVLIGDIWKQFPHAEIQKFSISEGFRDENNTVFORRQHSVFQRL  
EQNRIGAGTVKPMIDARRIWASIGRWMSLADRTTYYGKEARITDPPFGKRLARLTS  
VDHDLFRRLCDI"  
complement(11450. .11818)  
/notes="unnamed protein product; gene\_id:MPE11.4  
unknown protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB01050.1"  
/db\_xref="GI:9279592"  
/translation="MANSLKKEERVEEDNGKSDCNKPKSTEVVTVTVEEVDPEPKI  
LRKHVATRVAKVGVAEGLPFSKKRKRSQNLGRNSLDCNVGRDGBFDEINRVGL  
QGGLDLNCKPEPDSVLSL"  
16981. .17979  
/notes="gene\_id:MPE11.5"  
/codon\_start=1

exon

CDS

complement(342. .764)  
/notes="CDS is reported in Acc# AP000599  
contains similarity to CHP-rich zinc finger protein  
gene id:K9122.5"  
/number=1  
/evidence=not\_experimental  
join(1705. .2463,2548. .2739,2862. .2949,3037. .3107,  
3190. .3273,3410. .3514)  
/notes="gb|AAD55139.1  
gene id:MPE11.1"  
/codon\_start=1  
/evidence=not\_experimental  
/product="dihydroliipoamide S-acetyltransferase"  
/protein\_id="BAB01047.1"  
/db\_xref="GI:9279589"  
/translation="MTVRSKIRIFMPALSSSTMTEGKIVSWIKTEGEKLAKGESVVVV  
ESDKADMETFDVGGVLAIVVEGETAPVGAIGLAEAEIEAEKKAASKSSSS  
VBAVVPSPPTVSSAPALAPAPVATVSDGPRKVTAPYAKLAKQKHQVIESVAG  
TGFGRTASDVTAAGIAPSKSIAPPPPPPTAKATNTNLPPLPDSIVPPTA  
MOSVAKNMIESLVPFRVGVVNTDALDYKVKPKGVTMTALLAKAAGMALAQH  
PVNASKDKGSFYSNINIAVAINGLIIPVLQADKLDLYLLSOKWELVGKA  
RSKLOPHYNSGTFTLSNLGMFQVDRFDAILPPGOGAIMAVGASKPTTVADKGFSS  
KQNTMLVNTADHRIYVGADLAFLQTFAKLIENPDSLIL"  
complement(4594. .5106)  
/notes="unnamed protein product; gene\_id:MPE11.2  
unknown protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB01048.1"  
/db\_xref="GI:9279590"  
/translation="MEDLEEERLTRDTSVGNKRVRGCLDLDSPVVKRLRDDILFDSSGL  
DPVSQDLDSVKMSFENELSTTTAALSGETQPDGLYLFASDDELGLPPLTPQTLL  
PSCSETVTELVRASDSSEVGELCGFEDHVTFFGCDLGDGLFFYFDGCLDSGDLF  
SWRPEFLPAE"  
complement(join(10090. .10408,10649. .10804,10916. .10962))  
/notes="unnamed protein product; gene\_id:MPE11.3  
similar to unknown protein  
sp|P42744  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB01049.1"  
/db\_xref="GI:9279591"  
/translation="MMPEKAKYDRQLMYTIQGTLEBASICLLNCGPIGNSALKNLVLG  
GVGSIITVSGSKVLIGDIWKQFPHAEIQKFSISEGFRDENNTVFORRQHSVFQRL  
EQNRIGAGTVKPMIDARRIWASIGRWMSLADRTTYYGKEARITDPPFGKRLARLTS  
VDHDLFRRLCDI"  
complement(11450. .11818)  
/notes="unnamed protein product; gene\_id:MPE11.4  
unknown protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB01050.1"  
/db\_xref="GI:9279592"  
/translation="MANSLKKEERVEEDNGKSDCNKPKSTEVVTVTVEEVDPEPKI  
LRKHVATRVAKVGVAEGLPFSKKRKRSQNLGRNSLDCNVGRDGBFDEINRVGL  
QGGLDLNCKPEPDSVLSL"  
16981. .17979  
/notes="gene\_id:MPE11.5"  
/codon\_start=1

CDS

complement(342. .764)  
/notes="CDS is reported in Acc# AP000599  
contains similarity to CHP-rich zinc finger protein  
gene id:K9122.5"  
/number=1  
/evidence=not\_experimental  
join(1705. .2463,2548. .2739,2862. .2949,3037. .3107,  
3190. .3273,3410. .3514)  
/notes="gb|AAD55139.1  
gene id:MPE11.1"  
/codon\_start=1  
/evidence=not\_experimental  
/product="dihydroliipoamide S-acetyltransferase"  
/protein\_id="BAB01047.1"  
/db\_xref="GI:9279589"  
/translation="MTVRSKIRIFMPALSSSTMTEGKIVSWIKTEGEKLAKGESVVVV  
ESDKADMETFDVGGVLAIVVEGETAPVGAIGLAEAEIEAEKKAASKSSSS  
VBAVVPSPPTVSSAPALAPAPVATVSDGPRKVTAPYAKLAKQKHQVIESVAG  
TGFGRTASDVTAAGIAPSKSIAPPPPPPTAKATNTNLPPLPDSIVPPTA  
MOSVAKNMIESLVPFRVGVVNTDALDYKVKPKGVTMTALLAKAAGMALAQH  
PVNASKDKGSFYSNINIAVAINGLIIPVLQADKLDLYLLSOKWELVGKA  
RSKLOPHYNSGTFTLSNLGMFQVDRFDAILPPGOGAIMAVGASKPTTVADKGFSS  
KQNTMLVNTADHRIYVGADLAFLQTFAKLIENPDSLIL"  
complement(4594. .5106)  
/notes="unnamed protein product; gene\_id:MPE11.2  
unknown protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB01048.1"  
/db\_xref="GI:9279590"  
/translation="MEDLEEERLTRDTSVGNKRVRGCLDLDSPVVKRLRDDILFDSSGL  
DPVSQDLDSVKMSFENELSTTTAALSGETQPDGLYLFASDDELGLPPLTPQTLL  
PSCSETVTELVRASDSSEVGELCGFEDHVTFFGCDLGDGLFFYFDGCLDSGDLF  
SWRPEFLPAE"  
complement(join(10090. .10408,10649. .10804,10916. .10962))  
/notes="unnamed protein product; gene\_id:MPE11.3  
similar to unknown protein  
sp|P42744  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB01049.1"  
/db\_xref="GI:9279591"  
/translation="MMPEKAKYDRQLMYTIQGTLEBASICLLNCGPIGNSALKNLVLG  
GVGSIITVSGSKVLIGDIWKQFPHAEIQKFSISEGFRDENNTVFORRQHSVFQRL  
EQNRIGAGTVKPMIDARRIWASIGRWMSLADRTTYYGKEARITDPPFGKRLARLTS  
VDHDLFRRLCDI"  
complement(11450. .11818)  
/notes="unnamed protein product; gene\_id:MPE11.4  
unknown protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB01050.1"  
/db\_xref="GI:9279592"  
/translation="MANSLKKEERVEEDNGKSDCNKPKSTEVVTVTVEEVDPEPKI  
LRKHVATRVAKVGVAEGLPFSKKRKRSQNLGRNSLDCNVGRDGBFDEINRVGL  
QGGLDLNCKPEPDSVLSL"  
16981. .17979  
/notes="gene\_id:MPE11.5"  
/codon\_start=1

CDS

complement(342. .764)  
/notes="CDS is reported in Acc# AP000599  
contains similarity to CHP-rich zinc finger protein  
gene id:K9122.5"  
/number=1  
/evidence=not\_experimental  
join(1705. .2463,2548. .2739,2862. .2949,3037. .3107,  
3190. .3273,3410. .3514)  
/notes="gb|AAD55139.1  
gene id:MPE11.1"  
/codon\_start=1  
/evidence=not\_experimental  
/product="dihydroliipoamide S-acetyltransferase"  
/protein\_id="BAB01047.1"  
/db\_xref="GI:9279589"  
/translation="MTVRSKIRIFMPALSSSTMTEGKIVSWIKTEGEKLAKGESVVVV  
ESDKADMETFDVGGVLAIVVEGETAPVGAIGLAEAEIEAEKKAASKSSSS  
VBAVVPSPPTVSSAPALAPAPVATVSDGPRKVTAPYAKLAKQKHQVIESVAG  
TGFGRTASDVTAAGIAPSKSIAPPPPPPTAKATNTNLPPLPDSIVPPTA  
MOSVAKNMIESLVPFRVGVVNTDALDYKVKPKGVTMTALLAKAAGMALAQH  
PVNASKDKGSFYSNINIAVAINGLIIPVLQADKLDLYLLSOKWELVGKA  
RSKLOPHYNSGTFTLSNLGMFQVDRFDAILPPGOGAIMAVGASKPTTVADKGFSS  
KQNTMLVNTADHRIYVGADLAFLQTFAKLIENPDSLIL"  
complement(4594. .5106)  
/notes="unnamed protein product; gene\_id:MPE11.2  
unknown protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB01048.1"  
/db\_xref="GI:9279590"  
/translation="MEDLEEERLTRDTSVGNKRVRGCLDLDSPVVKRLRDDILFDSSGL  
DPVSQDLDSVKMSFENELSTTTAALSGETQPDGLYLFASDDELGLPPLTPQTLL  
PSCSETVTELVRASDSSEVGELCGFEDHVTFFGCDLGDGLFFYFDGCLDSGDLF  
SWRPEFLPAE"  
complement(join(10090. .10408,10649. .10804,10916. .10962))  
/notes="unnamed protein product; gene\_id:MPE11.3  
similar to unknown protein  
sp|P42744  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB01049.1"  
/db\_xref="GI:9279591"  
/translation="MMPEKAKYDRQLMYTIQGTLEBASICLLNCGPIGNSALKNLVLG  
GVGSIITVSGSKVLIGDIWKQFPHAEIQKFSISEGFRDENNTVFORRQHSVFQRL  
EQNRIGAGTVKPMIDARRIWASIGRWMSLADRTTYYGKEARITDPPFGKRLARLTS  
VDHDLFRRLCDI"  
complement(11450. .11818)  
/notes="unnamed protein product; gene\_id:MPE11.4  
unknown protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB01050.1"  
/db\_xref="GI:9279592"  
/translation="MANSLKKEERVEEDNGKSDCNKPKSTEVVTVTVEEVDPEPKI  
LRKHVATRVAKVGVAEGLPFSKKRKRSQNLGRNSLDCNVGRDGBFDEINRVGL  
QGGLDLNCKPEPDSVLSL"  
16981. .17979  
/notes="gene\_id:MPE11.5"  
/codon\_start=1

CDS

complement(342. .764)  
/notes="CDS is reported in Acc# AP000599  
contains similarity to CHP-rich zinc finger protein  
gene id:K9122.5"  
/number=1  
/evidence=not\_experimental  
join(1705. .2463,2548. .2739,2862. .2949,3037. .3107,  
3190. .3273,3410. .3514)  
/notes="gb|AAD55139.1  
gene id:MPE11.1"  
/codon\_start=1  
/evidence=not\_experimental  
/product="dihydroliipoamide S-acetyltransferase"  
/protein\_id="BAB01047.1"  
/db\_xref="GI:9279589"  
/translation="MTVRSKIRIFMPALSSSTMTEGKIVSWIKTEGEKLAKGESVVVV  
ESDKADMETFDVGGVLAIVVEGETAPVGAIGLAEAEIEAEKKAASKSSSS  
VBAVVPSPPTVSSAPALAPAPVATVSDGPRKVTAPYAKLAKQKHQVIESVAG  
TGFGRTASDVTAAGIAPSKSIAPPPPPPTAKATNTNLPPLPDSIVPPTA  
MOSVAKNMIESLVPFRVGVVNTDALDYKVKPKGVTMTALLAKAAGMALAQH  
PVNASKDKGSFYSNINIAVAINGLIIPVLQADKLDLYLLSOKWELVGKA  
RSKLOPHYNSGTFTLSNLGMFQVDRFDAILPPGOGAIMAVGASKPTTVADKGFSS  
KQNTMLVNTADHRIYVGADLAFLQTFAKLIENPDSLIL"  
complement(4594. .5106)  
/notes="unnamed protein product; gene\_id:MPE11.2  
unknown protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB01048.1"  
/db\_xref="GI:9279590"  
/translation="MEDLEEERLTRDTSVGNKRVRGCLDLDSPVVKRLRDDILFDSSGL  
DPVSQDLDSVKMSFENELSTTTAALSGETQPDGLYLFASDDELGLPPLTPQTLL  
PSCSETVTELVRASDSSEVGELCGFEDHVTFFGCDLGDGLFFYFDGCLDSGDLF  
SWRPEFLPAE"  
complement(join(10090. .10408,10649. .10804,10916. .10962))  
/notes="unnamed protein product; gene\_id:MPE11.3  
similar to unknown protein  
sp|P42744  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB01049.1"  
/db\_xref="GI:9279591"  
/translation="MMPEKAKYDRQLMYTIQGTLEBASICLLNCGPIGNSALKNLVLG  
GVGSIITVSGSKVLIGDIWKQFPHAEIQKFSISEGFRDENNTVFORRQHSVFQRL  
EQNRIGAGTVKPMIDARRIWASIGRWMSLADRTTYYGKEARITDPPFGKRLARLTS  
VDHDLFRRLCDI"  
complement(11450. .11818)  
/notes="unnamed protein product; gene\_id:MPE11.4  
unknown protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB01050.1"  
/db\_xref="GI:9279592"  
/translation="MANSLKKEERVEEDNGKSDCNKPKSTEVVTVTVEEVDPEPKI  
LRKHVATRVAKVGVAEGLPFSKKRKRSQNLGRNSLDCNVGRDGBFDEINRVGL  
QGGLDLNCKPEPDSVLSL"  
16981. .17979  
/notes="gene\_id:MPE11.5"  
/codon\_start=1

CDS

complement(342. .764)  
/notes="CDS is reported in Acc# AP000599  
contains similarity to CHP-rich zinc finger protein  
gene id:K9122.5"  
/number=1  
/evidence=not\_experimental  
join(1705. .2463,2548. .2739,2862. .2949,3037. .3107,  
3190. .3273,3410. .3514)  
/notes="gb|AAD55139.1  
gene id:MPE11.1"  
/codon\_start=1  
/evidence=not\_experimental  
/product="dihydroliipoamide S-acetyltransferase"  
/protein\_id="BAB01047.1"  
/db\_xref="GI:9279589"  
/translation="MTVRSKIRIFMPALSSSTMTEGKIVSWIKTEGEKLAKGESVVVV  
ESDKADMETFDVGGVLAIVVEGETAPVGAIGLAEAEIEAEKKAASKSSSS  
VBAVVPSPPTVSSAPALAPAPVATVSDGPRKVTAPYAKLAKQKHQVIESVAG  
TGFGRTASDVTAAGIAPSKSIAPPPPPPTAKATNTNLPPLPDSIVPPTA  
MOSVAKNMIESLVPFRVGVVNTDALDYKVKPKGVTMTALLAKAAGMALAQH  
PVNASKDKGSFYSNINIAVAINGLIIPVLQADKLDLYLLSOKWELVGKA  
RSKLOPHYNSGTFTLSNLGMFQVDRFDAILPPGOGAIMAVGASKPTTVADKGFSS  
KQNTMLVNTADHRIYVGADLAFLQTFAKLIENPDSLIL"  
complement(4594. .5106)  
/notes="unnamed protein product; gene\_id:MPE11.2  
unknown protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB01048.1"  
/db\_xref="GI:9279590"  
/translation="MEDLEEERLTRDTSVGNKRVRGCLDLDSPVVKRLRDDILFDSSGL  
DPVSQDLDSVKMSFENELSTTTAALSGETQPDGLYLFASDDELGLPPLTPQTLL  
PSCSETVTELVRASDSSEVGELCGFEDHVTFFGCDLGDGLFFYFDGCLDSGDLF  
SWRPEFLPAE"  
complement(join(10090. .10408,10649. .10804,10916. .10962))  
/notes="unnamed protein product; gene\_id:MPE11.3  
similar to unknown protein  
sp|P42744  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB01049.1"  
/db\_xref="GI:9279591"  
/translation="MMPEKAKYDRQLMYTIQGTLEBASICLLNCGPIGNSALKNLVLG  
GVGSIITVSGSKVLIGDIWKQFPHAEIQKFSISEGFRDENNTVFORRQHSVFQRL  
EQNRIGAGTVKPMIDARRIWASIGRWMSLADRTTYYGKEARITDPPFGKRLARLTS  
VDHDLFRRLCDI"  
complement(11450. .11818)  
/notes="unnamed protein product; gene\_id:MPE11.4  
unknown protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB01050.1"  
/db\_xref="GI:9279592"  
/translation="MANSLKKEERVEEDNGKSDCNKPKSTEVVTVTVEEVDPEPKI  
LRKHVATRVAKVGVAEGLPFSKKRKRSQNLGRNSLDCNVGRDGBFDEINRVGL  
QGGLDLNCKPEPDSVLSL"  
16981. .17979  
/notes="gene\_id:MPE11.5"  
/codon\_start=1

/evidence=not\_experimental  
/product="Ap2 domain transcription factor-like protein"  
/protein\_id="BAB01051.1"  
/db\_xref="GI:9279593"  
/translation="MAERKKRSSIQTNKPNKMKKPPQLNHLPLGLSBDLTKMRKL  
FVNDPYATDYSSREERSQRKRYVCEIDLPFAQAATQAESSESYCCSENNNGYSK  
TKLSACSKVLBSKASPVVGRSSTTVSPKPVGRQKWKGAWEAIEHRPIKTVTLWTY  
ETLEQADAYATKLEFDALAAATSAASVLSNESGSMISASSESSIDDLKLVDSLD  
QOAGSKASDFDADQIPEMGCFIDDSFIPNACELDFLLTEENNQMDDYCGID  
DLDIIGLCDSGSELPDYDFSDVEIDGLIGTITIDKYAFVDHIAITTPPLAIACP"  
join(21893. .22063,22330. .22386,22578. .22783,22894. .23227,  
23311. .23382,23473. .23613)  
/note="gb|AAF23821.1  
gene\_id:MPE11.6"  
/codon\_start=1  
/evidence=not\_experimental  
/product="homocysteine S-methyltransferase AtHMT-1"  
/protein\_id="BAB01052.1"  
/db\_xref="GI:9279594"  
/translation="MWLEKKSALLEDLIKKCGCAVVDGGFATQLEIHGAAINPLWS  
AVSLKNPELIRKVMEXLEAGADIVVTSSYQATIPGLSRGLSEESLLOKSVEL  
AVEARDREKVKVSGHSGYNRAVLAAGISYGAYLADGSEYSGHYGENVSLDKLXDF  
HRRRLQVLVEAGPDLLEPETINKLEAQAQVLLBEEKVQIPAWICFTSVGGEKAPSG  
ESFECLEPLKSNNIYAVINCAPPQFTENLIRKFAKITKKAIVVYPSNGGEWDGKA  
KWLPSQCFGDDEPFMPATKWRDLGAKLIGGCCRTTPSTINALSRLKRR"  
complement(27277. .27552)  
/note="unnamed protein product; gene\_id:MPE11.7  
unknown protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB01053.1"  
/db\_xref="GI:9279595"  
/translation="MTHAREWSSSTTTLLMVLISYMLHLFCVYSRVGAIRIFPETP  
ASGKQEDLMKKYFGAGKFPVDSFVGKISGSEKRVSPSCDPLHN"  
37065. .31883  
/note="unnamed protein product; gb|AAF16548.1  
gene\_id:MPE11.8  
similar to unknown protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB01054.1"  
/db\_xref="GI:9279596"  
/translation="MPKREKERSVSLDKYKRSPLCCEASLALKPSEKOVKEWEARCP  
VCMERPHNGILLICSSYENGCRPYMCDTSHRSNCFDQFRKASKKPSLSLREBES  
NPTMEDVDSDTAVNLGEAASETIVVDLSDGEBEVEEVEEVEVEVEEVEEIV  
TTEDEQKNKPKQLTCPLCRHIKEWVWVKAACFMKXKSCCTCDPSGSYDLR  
KHARLLHPVRPSEADPRQRWRRLEROSLDLSTLQSSFGGDEISNDGFLPAD  
TLLLTVPFLIRVPERSSSGSSWSGTSRARTHTSGRRSSRSPASLWGESYEGNTGT  
SPRENNQSSDEQVSGTFRRSRRTWIDDDDEEPP"  
complement(join(32197. .32430,32525. .32676,32969. .33332,  
33404. .33487))  
/note="gene\_id:MPE11.9"  
/codon\_start=1  
/evidence=not\_experimental  
/product="chloroplast 50S ribosomal protein L15"  
/protein\_id="BAB01055.1"  
/db\_xref="GI:9279597"  
/translation="MATPLISNPLTSRHCYRLHLSSTSFKNVSVLGANPQSILSL  
KLNQTKRNQQQFARPLVVSOTATSAVVAERFDNLNGPDSGSKKOKRKGK  
TSAGOGASCGFGMRGOKSRSGPIMRGEGOTALYRLPKLGGTAGMWSGLPKYLP  
VNIKIETAGFQEGDEVSLETILKQGLINPSSRERKLPILKILGTGELSMLKTFKARAF  
STQAKEUEASGCTLTUVGRKKWVKPSVAKNQARAEYFAKKRAAAAESEPAAS  
A"  
complement(join(33961. .34021,34114. .34226,34391. .34482,  
34602. .34800))  
/note="unnamed protein product; gb|AAF26483.1  
gene\_id:MPE11.10  
similar to unknown protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB01056.1"  
/db\_xref="GI:9279598"  
/translation="MKVWMLIIDESNASVDLLIWALENQKDTIESSKVIFAKPQKNS

CDS  
FTPPTVSSVGAQIYPFPSPNSELIRIAOERNKMIKALILEKAKKICLNHGIIKAT  
FDDGPKOLIRKIIIOERNINLIVTSDQSKKCTQNTQCSLIVVVKRLKQD  
Join (36033. .36150.36243. .36283.36497. .36565. .36707,  
36797. .36871)  
/notes="unnamed protein product: contains similarity to  
transcription-associated zinc ribbon protein  
gene\_id:MP11.11"  
/codon\_start=1  
/evidence=not experimental  
/protein\_id="BAB01057.1"  
/db\_xref="GI:9279599"  
/translation="MEKSRSEPLFCNLGTMVLVSKTYAECPHCKTTRNAKDIIDK  
EIAVTSADIRRELGSIFGEKIQAEELPKIKKACEKQHPDELVYITRQKSADEG  
QTTYXTCNCAHRTFEG"  
Query Match 100.0%; Score 962; DB 8; Length 83650;  
Best Local Similarity 100.0%; Pred. No. 9.6e-140; Indels 0; Gaps 0;  
Matches 962; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGTGTTCTCAGAAATAGCAGCAATATTTATATAAGCATGCAATTTCTTTATAGATCG 60  
DB 12780 TGTGTTCTCAGAAATAGCAGCAATATTTATATAAGCATGCAATTTCTTTATAGATCG 12721  
QY 61 CGAAGTTTAAAAACATATAGAAATGTTTACATATATACATGGTTTTTATTCGATAACA 120  
DB 12720 CGAAGTTTAAAAACATATAGAAATGTTTACATATATACATGGTTTTTATTCGATAACA 12661  
QY 121 TGCAAAATATTTATTTTATTTTATGAGTTTATTTGATAGCATGACAAATATTAATATAT 180  
DB 12660 TGCAAAATATTTATTTTATTTTATGAGTTTATTTGATAGCATGACAAATATTAATATAT 12601  
QY 181 CAGTGTAAATACATGTTTGTTCCTTAAATACATGCAATTTTAAATTCAGACATTTGTTT 240  
DB 12600 CAGTGTAAATACATGTTTGTTCCTTAAATACATGCAATTTTAAATTCAGACATTTGTTT 12541  
QY 241 TAAATCAAATCTAATCTCTTATATCACACGCAATTCAGGAAATTCAGGTAAAGAA 300  
DB 12540 TAAATCAAATCTAATCTCTTATATCACACGCAATTCAGGAAATTCAGGTAAAGAA 12481  
QY 301 GAAATTAAGAAATGAGAGATAGAGATTTCTATGAAAGAAAGAGAGAAATCATGAG 360  
DB 12480 GAAATTAAGAAATGAGAGATAGAGATTTCTATGAAAGAAAGAGAGAAATCATGAG 12421  
QY 361 TGAACAAAATAAAGAGATATGATGATATATTTTATGAGAGTGGTGAAGATTTATTTAGG 420  
DB 12420 TGAACAAAATAAAGAGATATGATGATATATTTTATGAGAGTGGTGAAGATTTATTTAGG 12361  
QY 421 AGAGGAGAGAGAAATAGAAAAGAAATGACATGGTGAATCTGAAAGAGATGAATTTGTG 480  
DB 12360 AGAGGAGAGAGAAATAGAAAAGAAATGACATGGTGAATCTGAAAGAGATGAATTTGTG 12301  
QY 481 TTAAGATGAAGAGAGAAAGAGAACTCCATGGCTTAAAGTCTCGTAAAGAGATGAAAG 540  
DB 12300 TTAAGATGAAGAGAGAAAGAGAACTCCATGGCTTAAAGTCTCGTAAAGAGATGAAAG 12241  
QY 541 AAAACAAAGAGGAGAGAAAGAGAAAGCTTAAATAGACTAACTATTGCCAAAATTTTC 600  
DB 12240 AAAACAAAGAGGAGAGAAAGAGAAAGCTTAAATAGACTAACTATTGCCAAAATTTTC 12181  
QY 601 TGTAGCCGCAAAATACATATTTTGTGTCAGGTTATTTTGTATTTCTTTGAAGTCAAAAG 660  
DB 12180 TGTAGCCGCAAAATACATATTTTGTGTCAGGTTATTTTGTATTTCTTTGAAGTCAAAAG 12121  
QY 661 TTATTTCTTACATATCTCTTAAATATATAGCCGATACCAATTTTCCACACATGGACTTC 720  
DB 12120 TTATTTCTTACATATCTCTTAAATATATAGCCGATACCAATTTTCCACACATGGACTTC 12061  
QY 721 CTTTATTTCCAAAAGTCAATTAAGTGTGACGTGATGATCTTACGCTTTTAAACATCGCAT 780  
DB 12060 CTTTATTTCCAAAAGTCAATTAAGTGTGACGTGATGATCTTACGCTTTTAAACATCGCAT 12001  
QY 781 GATGATGTCATTAGCATCAATCTCCACCGTCCAAATTTATTTAGTTGTTGCAATATCGAC 840  
DB 12001 GATGATGTCATTAGCATCAATCTCCACCGTCCAAATTTATTTAGTTGTTGCAATATCGAC 840

DB 12000 GATGATGTCATTAGCATCAATCTCCACCGTCCAAATTTATTTAGTTGTTGCAATATCGAC 11941  
QY 841 CGTCTAAGTTTCCACACCGCGCTATAGAGTTTCAATATAAATTTTAGCAAAATAAAT 900  
DB 11940 CGTCTAAGTTTCCACACCGCGCTATAGAGTTTCAATATAAATTTTAGCAAAATAAAT 11881  
QY 901 CAGCAAAATAATTTTCTTTGACTAAGCTTTAAACGCGCTTAAACATTTTCTTCTCGCTA 960  
DB 11880 CAGCAAAATAATTTTCTTTGACTAAGCTTTAAACGCGCTTAAACATTTTCTTCTCGCTA 11821  
QY 961 AC 962  
DB 11820 AC 11819  
RESULT 2  
AR488147 LOCUS AR488147 1700 bp DNA linear PAT 15-MAY-2004  
DEFINITION Sequence 3 from patent US 6706952.  
ACCESSION AR488147  
VERSION AR488147.1 GI:47253921  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1700)  
AUTHORS Cad, R.M. and Dietrich, R.A.  
TITLE Arabidopsis gene encoding a protein involved in the regulation of  
SAR gene expression in plants  
JOURNAL Patent: US 6706952-A 3 16-MAR-2004;  
FEATURES  
Location/Qualifiers  
1..1700  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 89.6%; Score 862; DB 6; Length 1700;  
Best Local Similarity 100.0%; Pred. No. 6.8e-124;  
Matches 862; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 101 TGGGTTTTTATTTGGATAACATGACAAATATTTTATTTATTTTCATGATTTTATTTGGATAG 160  
DB 1 TGGGTTTTTATTTGGATAACATGACAAATATTTTATTTATTTTCATGATTTTATTTGGATAG 60  
QY 161 CATGACAAATATTAATATATATATATAGTGTGTTAATAACATGTTTGTCTTAAATACATCAT 220  
DB 61 CATGACAAATATTAATATATATATATAGTGTGTTAATAACATGTTTGTCTTAAATACATCAT 120  
QY 221 TTTAAATCAGACATTTGTTTTTAAATCAAATCTAATCTCTTATATCAACGACATTGAC 280  
DB 121 TTTAAATCAGACATTTGTTTTTAAATCAAATCTAATCTCTTATATCAACGACATTGAC 180  
QY 281 GGAAATTCAGGTAAAGAGAAATATAAGATGAGAGATAGAGATTTCTATGAAATA 340  
DB 181 GGAAATTCAGGTAAAGAGAAATATAAGATGAGAGATAGAGATTTCTATGAAATA 240  
QY 341 AGAAGAGAGAAATAGTGTGTAACAAATAAAGAGATATGATATATTTTATGAGAG 400  
DB 241 AGAAGAGAGAAATAGTGTGTAACAAATAAAGAGATATGATATATTTTATGAGAG 300  
QY 401 GTGTGGAAGATTTATTTTAGGAGGGGAGAGAGAAATAGAAAAGAAAATGACATGTTGAA 460  
DB 301 GTGTGGAAGATTTATTTAGGAGGGGAGAGAGAAATAGAAAAGAAAATGACATGTTGAA 360  
QY 461 TCTGAAGAGAGATGAATTTGTTTAAAGATGAAGAGAGAAAGAACTTCCATGGCTTAAAGTC 520  
DB 361 TCTGAAGAGAGATGAATTTGTTTAAAGATGAAGAGAGAAAGAACTTCCATGGCTTAAAGTC 420  
QY 521 TCGTAAAGAGATGAAAAGAAACAAAGAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAG 580  
DB 421 TCGTAAAGAGATGAAAAGAAACAAAGAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAG 480  
QY 581 CTAACCTATTGCCAAAATTTCTGTAGCCGACAAATACTATTTTGTGTCACAGGTTATTTTGTG 640

Db	481	CTAACTATTGCGCAAAATTTCTGTAGCGGCAAAATACTATTTTGGTCCAAAGTTATTTTGTG	540
Qy	641	TATCTTTTGAAGTCAAAAGTTATTTCTTACATATACTCTAAATAATATAGCCGATACCAA	700
Db	541	TATCTTTTGAAGTCAAAAGTTATTTCTTACATATACTCTAAATAATATAGCCGATACCAA	600
Qy	701	TTTTTCCACATGGAAGTCTCTTTTATTCACAAAGTCAATAAGTGTGACGTCAATGATACT	760
Db	601	TTTTTCCACATGGAAGTCTCTTTTATTCACAAAGTCAATAAGTGTGACGTCAATGATACT	660
Qy	761	TAGCGTTTAAACATCGCATGATGATGTCAATTAGCATCAATCTCCACCGTCCCAATTTATT	820
Db	661	TAGCGTTTAAACATCGCATGATGATGTCAATTAGCATCAATCTCCACCGTCCCAATTTATT	720
Qy	821	TAGTTGTTGACATATCGACCGTCTAAGTTTCCACACCGCGGTATAAGAGTTTTCATTAT	880
Db	721	TAGTTGTTGACATATCGACCGTCTAAGTTTCCACACCGCGGTATAAGAGTTTTCATTAT	780
Qy	881	AAATTTTAGCAAAATAAATCAGCAATAATATTTTCTTGACTAAGCTTAAACGACGCCG	940
Db	781	AAATTTTAGCAAAATAAATCAGCAATAATATTTTCTTGACTAAGCTTAAACGACGCCG	840
Qy	941	TTAAACATTTTCTCTCGCTAAC	962
Db	841	TTAAACATTTTCTCTCGCTAAC	862
RESULT 3			
AB026636			
LOCUS			
DEFINITION			
AB026636			
VERSION			
AB026636.1			
KEYWORDS			
SOURCE			
ORGANISM			
Arabidopsis thaliana (thale cress)			
Arabidopsis thaliana			
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eutrosia II; Brassicales; Brassicaceae; Arabidopsis.			
1			
Sato, S., Nakamura, Y., Kaneko, T., Katoh, T., Asamizu, E. and Tabata, S.			
Structural analysis of Arabidopsis thaliana chromosome 3. I.			
Sequence features of the regions of 4,504,864 bp covered by sixty			
P1 and TAC clones			
DNA Res. 7 (2), 131-135 (2000)			
20277480			
10819329			
REFERENCE			
2 (bases 1 to 92620)			
Sato, S., Nakamura, Y., Kaneko, T., Kato, T., Asamizu, E. and Tabata, S.			
Direct Submission			
Submitted (28-APR-1999) Yasukazu Nakamura, Kazusa DNA Research			
Institute, Department of Plant Gene Research, 1532-3, Yana,			
Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp,			
Tel: 81-438-52-3935, Fax: 81-438-52-3934)			
Address for correspondence: kaos@kazusa.or.jp			
For the latest information on annotation of this clone, please see			
http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=K14A17			
Genes with similarity to proteins in the databases are described in			
'product' or 'note' qualifiers. Genes that have no significant			
protein similarity are described as 'unknown protein'			
The software programs used to predict genes include: Grail			
(Informatics Group, Oak Ridge National Laboratory,			
http://compbio.ornl.gov/Grail-1.3/),			
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),			
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of			
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and			
SplicePredictor (Volker Brendel, Stanford University,			
http://gremli.zool.iastate.edu/cgi-bin/sp.cgi).			
Genes encoding tRNAs are predicted by tRNAscan-SE			
(Sean Eddy, Washington University School of Medicine, St. Louis,			
http://genome.wustl.edu/eddy/tRNAscan-SE/).			
This sequence may not be the entire insert of this clone. It may be			

shorter because we remove overlaps between neighboring submissions. The 5' clone is MUI15 and the 3' clone is MCE21.	
FEATURES	
Location/Qualifiers	source
1..92620	
/organism="Arabidopsis thaliana"	
/mol_type="genomic DNA"	
/db_xref="taxon:3702"	
/chromosome="3"	
/clone_lib="K14A17"	
/ecotype="Columbia"	
/complement(201..2180)	
/note="gene_id:K14A17.1"	
/codon_start=1	
/evidence=not_experimental	
/product="salt-inducible protein-like"	
/protein_id="BAA94973.1"	
/db_xref="GI:7670019"	
/translation="MRGFASASRIATAAAASKSLNASTSVNPKLSKTLNSSGKPTNP LNQYISQVIERKDWFLILNQEFTHRIGLNTFRFVSVLQNDQNPUSRFYLDWNSF DPYAKDSLSKLGNALFRKGPLLSMELKEIRDSYRISDELMCVLIGWGRGLG ARYCNDVFAQISFLGMPSTRLYNAIDALVKSNSLDLAYLKFQPMRSGCCPKDRFTY NLIHGVCCKGVDEAIRLVKOMEQGNRPVFTVILIDGLIAGRVDALVCLSNNS RVKLNPEATIRTFVHGI FRC LPPCKAFVLVGFMEKDSNLQVGYDAVLVCLSNNS MAKETGQFLKIGERGYIPDSSTFNAAMSLCKGHDLVETCRI FDGFSVRGKVPGRNG YLVVQALINARQFSEGRYLRKQMGVGLSSVSYNAVIDCLCKARRIENAMFLUTE MODRISPNLVFTFSLGYSVRGDKVHGLEKLVHGFDPKVTFFLLIINCLCRA KEIKDAFCFEMLEWGEIPNEITINILIRSCCTGDTDSVKLFAMKENGSLSPDY AYNATIOSFCMKRVKAEELKTLRIGLKPDPNTYSTLIKALSESRESEAREMFS STERHGCVPDSYTKRLVELDLRKSLGSLRETVSAS"	
join(3906..4124,4238..4462,4536..4649)	
/note="unnamed protein product; emb CAB82745.1	
gene_id:K14A17.2	
similar to unknown protein"	
/codon_start=1	
/evidence=not_experimental	
/protein_id="BAA94974.1"	
/db_xref="GI:7670020"	
/translation="MAKQDPTSSNTMLPLVGESEFVRPQPQLDITITGDTVKDATGNKVF KYKTLFGLHNKRILVDPNDSIPTVMKMKVTSKIDRWQVYRGSDDDKIFTKRSSTV QLKTRVEFLKHNQTRSCSDFTIKGRFMKRACTIYVADSTKIIAQVYEGHERLVATI VNPVYAFITVLI FIFDILINMGTTGI"	
complement(join(5031..5927,6011..6601,6984..7211))	
/note="gene_id:K14A17.3"	
/codon_start=1	
/evidence=not_experimental	
/product="AMP-binding protein"	
/protein_id="BAA94975.1"	
/db_xref="GI:7670021"	
/translation="MAATKWRDIDDLPKIPANYTALTPLWFLDRAAVVHPTRKSVIHG SREYTRQTYDRCLASALADRSIGPGSTVFVAI IAPNIPAMYEAHFVPMCGAVLN CVNIRLNAPTVALLSHSQSSVIMVDQFEFTLAEDSLRLMEEKAGSFKRPLLIVIGD HTCAPESLNRLSKGAIEYEDFLATDPDNPWPADQEWSTALGYTSTGTTASPKGV LHRRGALMALSNPLIWMQDCAVYLTLPMPHCNGWCFFWSLAVLSGTSICLRQVTA KEVYSNIAKYKTHFCAPVNLAINAPKEDTILPLHTVHVMTAGAAPPSVLSFM NQGRFPAVYGLSETYGPSTCAWKPEWSDLPPTQAKLNARQVRYTMEQLDVID TQTGKRPADGKTAGEIVFRGNMVKGLKNPEANKETFAGGWSHGDIAVXHPDNYI ELKRSQDVIIISGGENISVEVENYVYHHPVLEASVVARPDERMSPCAFTVLSKD YKHPQNKLAQDIMKFCREKLPAYVWPVKSVEFGPLPKTATGKTIQKHILRTAKEMGPV PESRU"	
complement(join(8463..8892,8977..9136,9230..9641))	
/note="gene_id:K14A17.4"	
/codon_start=1	
/evidence=not_experimental	
/product="basic chitinase"	
/protein_id="BAA94976.1"	
/db_xref="GI:7670022"	
/translation="MVSRLPFSLLLTVALVVFQGTLVNAEDSEPSSTRKPLVKIV KKGKICDGMGECKGWSYCCNHTIISDFETQYFENLFSKNSPFAHGVFWDYRSFIT AAAYQPLGFTAGKLGQMEKVAAPLGHVSGTSCGYGATGGLAWGLCYKNKMS DQLYCDDYVKLTYPCTPGVSHRGALPVYNNYNYQGTGALKVLLSHPEVYLNENAT LAFQAAIWRWMTPPKKHLPASHDFVGWKPKNDTAAKRTPGFGATINVLYGDOI CN SGFDNEMNIVSHVLYYLDLIGVGREAGPHEKLSQADQEPSPSSSSSAPPSSGSSS"	

```
CDS
join(13759..13822,14371..14445,14592..14707,14800..14888,
15014..15059,15238..15433,15520..15845,15928..17178,
17360..17713)
/notes="gene_id:K14A17.5"
/codon_start=1
/evidence=not_experimental
/product="transcription factor-like protein"
/protein_id="BAA94977.1"
/db_xref="GI:7670023"
/translations="MDGDLGRIGSEIHGFHTLQDLVDVTMLEEAKSRWLNRPEIHA
ILCGRIILFDKMLRPNKRDGHNKKKDGRTVKEAHEHLKVGNEIRIHVYVAGEDN
TTFVRCYLLDXARENILVLYHYRDQEAATTSQDSISSEISVSEQTFFNRYVAEDID
TVNRHDIISLHNDLNDLDELLVTDNLOSAPTVDNLSYTFEPLQNAANGTAHEGNA
TVADGSLDALNDPOSRESFGWMSNFISESGSLSDPSFEPVMVPRODPLAPQAVF
HSHSNTPEQVNTIDVSPAWAYSEKTKILVTGFLHDSYQHLERSNLYCVCGDFCPVA
EYLQAGVYRLLIIPPHSGVMNLYLSADGHPKIQCPRFEHRAVPVLDKTVPEONQPSK
WEFEQVYRLLIIFTSNKLNLSSKISPHNLDKAKLASKTNHLLNSWYLVKSIQ
GNKVSFQADKHLFELSLKRLKEWLMKEVLEGNTLDVDSKGLVHLCASLGYSIWS
VQULFSGLSNFRDKQGTALHWAAYVGRKVAALLSAGARPNIIVTDSTKDNLGGC
MAADLAQONGYDGLAAYLAECILVAQFRDMKIAGNITGDLEACKEMLNQGTLPDEQC
SLKDAAAYRTAAEAARAIQCAFREKALKAAARSVIOFANKKEEAKSIITAAMKIQNAF
RKYDTRRKIEAAVRIOCRPOTKIRREYLNMRRQAIRIOAARFGLQARQYKILWSV
GVLEKAVLRWKEKGFGLQVAAEDSPGEAQEDFYKTSQRAERLERSVVRVQAM
FRSKKAQDYRRMKLTHEBAQVNHILFLNLSFGKNSNRK"
complement(join(18170..18172,18407..18490,18736..18807,
18977..19090,19168..19281,19387..19487,19590..19668,
19755..19805,19906..19989,20365..20431,20539..20669,
20916..21040,21126..21231,21349..21483,21624..21770,
21860..22171))
/notes="gene_id:K14A17.6"
/pseudo
/codon_start=1
/evidence=not_experimental
/product="lipamide dehydrogenase"
join(29096..29206,29695..29826,29925..30009,30149..30312,
30830..31267)
/notes="unnamed protein product; contains similarity to
similar to ubiquitin conjugating enzyme
gene_id:K14A17.7"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA94978.1"
/db_xref="GI:7670024"
/translations="MADERYRNKNPNVAKRILQVKEQMANPNSDDFMSLPLEENIFEWQ
FAJRGDGFEGGIVHGRIQLPADVPKPPSPMLTPNGRPTNTKICLSISNYHPE
HMQPSVSTALVALLAFMPTSPNGALGSVDVPEKERRTLAKRSRTPKYGSPERQK
IIDEIHQYLSKATVPKPELPCSQAPSIYSAHQVPEQEAITVVERSTATTDTI
VDQIITEETAENVTRAASVVPAAAPLPAVEVVVKSVSQGEQRMARAAOKPVDVDRIFT
WAAVGUTIAIMVLLKKFKNSGYSTGFMDDQS"
join(32260..32364,33121..33412,33506..33750,33837..34103)
/notes="unnamed protein product; gb|AAD26962.1
gene_id:K14A17.8
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA94979.1"
/db_xref="GI:7670025"
/translations="MCKSSNIVEDRIKEBQMLSFKEIKQADLSDDCMRALPYSFIS
KVEKDFSYKMWIRAQGWKTNDVEVSKNFTYIETRGWDLFVSDNALGONEFITFTH
RGNMVFHVIYEQNGLEMRRKPRFQTPGSSGIKKEGENSELIDVKEESESPOGRA
EFLVRKKTKEDSKSSKTRNKRKKSQSKQVLDGVPEFKITIRKSYLFLAIPK
HFVDDHI PNKSKITFIRHPNGSGMWKLVCLVREIRTFISGYGSKLAREFFLWVGDKCTF
KLIKPFPEVLLTSSKKNREKMDQCMID"
complement(join(34528..34725,35213..35322,35680..35863))
/notes="unnamed protein product; gb|AAF26101.1
gene_id:K14A17.9
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA94980.1"
/db_xref="GI:7670026"
/translations="MAESGGRRIGVAVDFSDCKKALSWAIDNVVRDGHLLITIAH
DMNYERGEQJMETVGSFPFIPMSFEADAAVMKKYALKPDAETLIDIVNTAARKKTIW
```

```

MKIYWDGPPREKICAAAEQIPLSSLVGNRGLGLKRMIMGSVSNHVNVNACPVTVVK
AHT"
complement(join(36309..36416,36556..36846,37250..37324,
37500..37634,38020..38145,38221..38374,38433..38573,
38878..38987,39134..39281,39474..39751,39993..40262))
/notes="unnamed protein product; gene_id:K14A17.10
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA94981.1"
/db_xref="GI:7670027"

Query Match 70.3%; Score 676.2; DB 8; Length 92620;
Best Local Similarity 99.6%; Pred. No. 1.2e-95;
Matches 678; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGTGTTTTCAGAAAATAGCAGCAAAATATTTTATAAAGCATGCAATTCCTTATAGATCG 60
Db 71376 TGTGTTTTCAGAAAATAGCAGCAAAATATTTTATAAAGCATGCAATTCCTTATAGATCG 71435

QY 61 CGAAGTTTAAAAAACAATATAGAAATTTTACAATATTTACATGGTGTTCCTTATGATAACA 120
Db 71436 CGAAGTTTAAAAAACAATATAGAAATTTTACAATATTTACATGGTGTTCCTTATGATAACA 71495

QY 121 TGACAAATATTTATTTTTCATGAGTTCCTTATGATAGCATGACAAATATTAATATAT 180
Db 71496 TGACAAATATTTATTTTTCATGAGTTCCTTATGATAGCATGACAAATATTAATATAT 71555

QY 181 CAGTGTTAATACATGTTTGTTCCTTAAATACATGCAATTTTAAAAATCAGACATTTGTTT 240
Db 71556 CAGTGTTAATACATGTTTGTTCCTTAAATACATGCAATTTTAAAAATCAGACATTTGTTT 71615

QY 241 TAAATCAATCTAATCTCTTATATCACACGACATTTGACGGAATAATTCAGGTAAAAAGA 300
Db 71616 TAAATCAATCTAATCTCTTATATCACACGACATTTGACGGAATAATTCAGGTAAAAAGA 71675

QY 301 GAAATAAAGAAATGAGAGATAGAGAGATTTCTATGGAATAAAGAGAGAAACATGTAGG 360
Db 71676 GAAATAAAGAAATGAGAGATAGAGAGATTTCTATGGAATAAAGAGAGAAACATGTAGG 71735

QY 361 TGACAAATTAAGAGATATGATATATATTTTATGAGAGGTGGTGAAGATTAATTTTAGG 420
Db 71736 TGACAAATTAAGAGATATGATATATATTTTATGAGAGGTGGTGAAGATTAATTTTAGG 71795

QY 421 AGAGGGAGAGAGAAATAGAAAAAGAAATACATGCTGTAATCTGAAGACATGAATTTGTG 480
Db 71796 AGAGGGAGAGAGAAATAGAAAAAGAAATACATGCTGTAATCTGAAGACATGAATTTGTG 71855

QY 481 TTAAGATGAAGAGAGAAAGAGAACTCCATGGCTAAAGTCTCGTAAAGAGATGAAAAAG 540
Db 71856 TTAAGATGAAGAGAGAAAGAGAACTCCATGGCTAAAGTCTCGTAAAGAGATGAAAAAG 71915

QY 541 AAACAAAGAGAAAGAGAAAGAGAAAGCTTAAATAGACTAACTATTTCGCCAAAAATTC 600
Db 71916 AAACAAAGAGAAAGAGAAAGAGAAAGCTTAAATAGACTAACTAACTATTTCGCCAAAAATTC 71975

QY 601 TGTAGCCGCAATACTATTTCGTCCAGGTATTTTGTCTTCTTTTGAAGTCAAAAG 660
Db 71976 TGTAGCCGCAATACTATTTCGTCCAGGTATTTTGTCTTCTTTTGAAGTCAAAAG 72035

QY 661 TTATTTTCTTACATATACTCTA 681
Db 72036 TTATTTTCTTACATATCTCA 72056

RESULT 4
AB005248/c
LOCUS
Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MX110.
DEFINITION
AB005248 BAA00015
ACCESSION
AB005248.1 GI:2264320
VERSION
KEYWORDS
Arabidopsis thaliana (thale cress)
SOURCE
```

ORGANISM	Arabidopsis thaliana	/note="unnamed protein product; gene_id:MX110.2
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	pir  T03994
AUTHORS	1 Sato, S., Kotani, H., Nakamura, Y., Kaneko, T., Asamizu, E., Fukami, M., Miyajima, N. and Tabata, S.	/codon_start=1
TITLE	Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence features of the 1.6 Mb regions covered by twenty physically assigned P1 clones	/evidence=not experimental
JOURNAL	DNA Res. 4 (3), 215-230 (1997)	/protein_id="BAB09347.1"
MEDLINE	97471969	/db_xref="GI:9758813"
PUBMED	9330910	/translation="MWASSLSPFPCLSYFKMPAIALNSPRTTTPYNLKKQTRLHLKPLSGLMKEVIEQRKSERENPPLVFGHSGYAAWAEKNLWPFSSGFGDSYAVSLQGSDEPLGTAGTLOTTHASDIADFIENLGSPPVLVGHSGFGLIVQYLANIVNKRSLGTENAFPELSGAVMVCSVPPSGNSGLVRLYFSKPVAAFKVTLSLAAGFQKSIPLCRETFSQAMDDOLVKRYNLAR"
REFERENCE	2 (bases 1 to 83646)	join(7788..7835,8043..8777,9032..9157,9205..9901,10085..12958)
AUTHORS	Nakamura, Y.	/note="gene_id:MX110.3"
TITLE	Direct Submission	/pseudo
JOURNAL	Submitted (02-JUL-1997) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research, 1532-3, Yana, Kisarazu, Chiba 252-0812, Japan (E-mail:ynakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)	/codon_start=1
COMMENT	Address for correspondence: kaos@kazusa.or.jp For the latest information on annotation of this clone, please see <a href="http://www.kazusa.or.jp/kaos/cgi-bin/aggd_graph.cgi?c=MX110">http://www.kazusa.or.jp/kaos/cgi-bin/aggd_graph.cgi?c=MX110</a> Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, <a href="http://combio.ornl.gov/Grail-1.3/">http://combio.ornl.gov/Grail-1.3/</a> ), GENSCAN (Chris Burge, MIT, <a href="http://CCR-081.mit.edu/GENSCAN.html">http://CCR-081.mit.edu/GENSCAN.html</a> ), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <a href="http://www.cbs.dtu.dk/services/NetGene2/">http://www.cbs.dtu.dk/services/NetGene2/</a> ) and SplicePredictor (Volker Brendel, Stanford University, <a href="http://grenlin.zool.iastate.edu/cgi-bin/sp.cgi">http://grenlin.zool.iastate.edu/cgi-bin/sp.cgi</a> ). Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, <a href="http://genome.wustl.edu/eddy/tRNAscan-SE/">http://genome.wustl.edu/eddy/tRNAscan-SE/</a> ). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is MG17 and the 3' clone is MB18. Location/Qualifiers 1..83646 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /db_xref="taxon:3702" /chromosome="5" /clone="MX110" /clone_lib="Mitsui P1" /ecotype="Columbia" join(1791..2912,3016..3318,3584..4660) /notes="gene_id:MX110.1" /codon_start=1 /evidence=not experimental /product="disease resistance protein-like" /protein_id="BAB09346.1" /db_xref="GI:9758812" /translation="MTEKATDVSNMNNYSPSRDPDGLGMDAHKEMESLLCLDSD EYRMIGWPGIGKTIARVLYSFSNFELSIIFMGNIKELMYTRPVCSDYSAKIQ LOKQFLSQINHKDMELHGLVAQRLNDKKVLI VLDSDIQSLDADAI AKETRFWGHG SRITITPDQKLKANGINHVIKYVEPPSAVEAQMFCMAFGONPNDGFEELAWEVTKLLHPLGLRVGSHFGMRSHEWYNALPRLKIRLDASIQILKFSYDALCEEKDL FHLHACLFNNQNEVEDYALSFDRQGFHLLAEKSLINLKFSTNCTRIEMNLL VOLGDIIVRHKGHQSICPEGRKQFLIDARDICEVLTDTNGTNRNVVIGIFLEVRNLSCQ LNIISERAFQGNLKLRFPHDYDESXLYLPQGLNNLPQKLRIEWFRRPMTCLPS NFCTKYLIVRMKNSKLQWQGNLPGLNKRMIDLSESKHLKELPDLSTANLEYIM SCGISLVEPSSISGKLKRLMLSLRCQSKLEALPTNINLESYLDLDTCLLKKPPE ISTNIKDLKLTAKIYVSTTIKWSHLKLEMSYSENKELPHALDITITTYINDTE MEIIPQWKVISHLQTLGLEGCKRLVIFQLSDLSQLVVTYNGSSIMVNLNRPPLST LRFLWNLNCFKLNNEAREFIIOTSHALPSREVPANFTYRANGSSIMVNLNRPPLST TRFQACLVATSKKIDNDKEAADRTTVIPRIENDKIGVDVPMRYRPHVPTTLEOHLH TTFQIEADVTSNKLIFSDFIKDNNQAVIKESGVQLQ"	/complement(13525..13827) pir  T02618 similar to unknown protein" /codon_start=1 /evidence=not experimental /protein_id="BAB09348.1" /db_xref="GI:9758814" /translation="MKDVLIVTYSLIKMTKLSRCPDLCNSETYFOLDKHLKNEHDG LVLLMVDPSQWRCEQMKRHYEVDLMATEPPIAEVYVHOLPYDITLIIERCLSS join(18792..18841,19220..19412,19494..19579,19875..19959,20464..20607,20697..20783,20894..21006,21097..21162,21216..21308,21350..21509) /note="unnamed protein product; gene_id:MX110.5 unknown protein" /codon_start=1 /evidence=not experimental /protein_id="BAB09349.1" /db_xref="GI:9758815" /translation="NMSPKQISDDRGSSHFHRTFPQIIHLIGNFIRMSVSMYRYLN QTGAPVLLFCFLVPSIIIFLIIOKPMKGRALSNOQIVPSLINGVITAYLFIWKGK LKSCGPLRALILSEYSGVLGVLGYGRGHVWKLPKTDSTETKEERVQTEQALGM MQMMPVPAIGILSALRVIARRVSLKNQOKRLHAITITSATCFLPVPMWMLIIGSS SGTSELFPSSAWAFUSTIIFGIILIFYDNIABERLHMVFSFPHLMVAGVCLIVMDI RVNIRGNISREQLERLFEVTRRSIKRNPSCSIFPYTNDVLLNLVLDVDDVNLSGFIR FSIYPSPFSFPCFLGEAYANILV" join(26073..30268,30416..31552) /note="gene_id:MX110.6" /pseudo /codon_start=1 /evidence=not experimental /product="retroelement pol polyprotein-like" 32202..33620 /note="unnamed protein product; contains similarity to retroelement pol polyprotein gene_id:MX110.7" /codon_start=1 /evidence=not experimental /protein_id="BAB09350.1" /db_xref="GI:9758816" /translation="WVEGTRTNPKRSTWSTEQSPSEPVAAHTWKEDKALKVISPPLP SPPTFANREILKRGLETTIDEFFTKMGLGFTSMOHPTEPTVFSTWETPTFN PIKPLAKGGINFKVTKAFSIPDLCEAYGENKSEMSFPKGIHFHFWGELASGK FSCNIAIKSVRHPVIRYALKLAHALFQGETSSTTSEMCFLFQGVKELLVEDADEN GNAFLRVHDDLDGNNVGCIFASYLEDYKNRLIKGTYIIGTITPLPFAVKAGVDSLSP FNALPVKEDYDENLVRSCSLKRSDDAFILVTDREDNKLKCLPSKETTNVESHEDI MFLPSKEVQVPLEQFMEIEEDLPPSRDDDPYNLKNCDLPSTFPVPTKMEKFLOS VIKSQIKNKWSKSYAIGKLLKVKVQLQPPDYVSEDELVPSPNEEDRDNDILSDENNE TPATSSHSRKNRNASHSYMP"
CDS		join(36310..37248,37409..37549,37633..37824,38363..38519) /note="emb CAB86928.1 gene_id:MX110.8 similar to unknown protein"



```
/pseudo
/codon_start=1
/evidence=not experimental
join(39564, 40460, 40627, 40767, 40853, .41185)
/note="unnamed protein product; gb|AAD55600.1
gene_id:MX110.9
strong similarity to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAB09351.1"
/db_xref="GI:9758817"
/translation="MDLSNLPYELLCILHSFLITTEAALTSVLSKRRNRLIAFVNV
DIVDHIREFLMDFVRVLALQGNPINKFSLDCSGVSDRVCNIQNVVRGVSELN
LSFVDSFEDDNLFPKVFENKVLKGLSYISWLDGSIFFLMLKTLVLESVLVSVE
KPFILHLPALBELVNNIYWKERDVVDLIVSSESLKTLTINFI VCTHTLSFDPIL
AYLSYGVVVDVVEAKMENLPEARISLVVEGDISRVRALINNDLLEDDEVDVLOPE
NVWLKMGIRNIRCLYLSNPTLEVLISLCCESMPVFNKLSLSIKSARENRQWAMPVLL
RNCPHLETVLVLEGLLHHVTDKGDACDCVSRDKRSLTSCPVKVLIEIKGFQGTIKEM
NMIKHLEYFPFCLKELKMYMEENDPAQLRVPEVITGEMMERYKKSSSCNVQLLVGGL
YKXWTP"
join(42817, 43766, 43839, .43979, 44083, .44403)
/note="emb|CAB86928.1
gene_id:MX110.10
similar to unknown protein"
/pseudo
/codon_start=1
/evidence=not experimental
join(45089, 45677, 46843, 47158, 47312, 47452, 47544, .47871)
/note="contains similarity to non-LTR retroelement reverse
transcriptase
gene_id:MX110.11"
/pseudo
/codon_start=1
/evidence=not experimental
/protein_id="BAB09352.1"
/db_xref="GI:9758818"
/translation="MIVDSMKRVARVAVVEGKERNSSNGEETRRSOSKHARTAKAR
MNVYKESKLPKPKKSLMIISLWKKEDTODNNVTVEKEKEKEKGTQVNTIRE
NAELMSKCVPEPHMLTYEEETWFAPFGWQWORYDAVMSGKSWFN"
complement(join(50633, .50872, 51043, .51177, 51268, .51438))
/note="gene_id:MX110.13
sp|P10798"
/codon_start=1
Query Match 65.7%; Score 632.2; DB 8; Length 83646;
Best Local Similarity 93.8%; Pred. No. 7.8e-89;
Matches 669; Conservative 0; Mismatches 43; Indels 1; Gaps 1;
QY 1 TGTGTTTCTCAGAAATAGCAGAAATATTATATAAAGCATGCAATCTCTTATAGATCG 60
Db 23871 TGTGTTTCTCAGAAATAGCAGAAATATTATATAAAGCATGCAATCTCTTATAGATCG 23812
QY 61 CGAAGTTTAAATAACATATAGATTTGTTTACATATATACATGGTTTATTCGATAACA 120
Db 23811 CGAAGTTTAAATAACATATAGATTTGTTTACATATATACATGGTTTATTCGATAACA 23752
QY 121 TGACAAATATTTATTTTATTTTCATGAGTTTATTTTGGATAGCATGACAAATATTAAATATAT 180
Db 23751 TGACAAATATTTATTTTATTTTCATGAGTTTATTTTGGGATAGCATGACAAATATTAAATATAT 23692
QY 181 CAGTGTATAACATGTTTGTGTTTCTTAAATAACATGCAATTTAAATTCAGACATTTGTTT 240
Db 23691 CAGTGTATAACATGTTTGTGTTTCTTAAATAACATGCAATTTAAATTCAGACATTTGTTT 23632
QY 241 TAAATCAATCTAAATCTCTTATATCACACGCAATTCAGCGAAATTCAGGTAAAGA 300
Db 23631 TAAATCAATCTAAATCTCTTATATCACACGCAATTCAGCGAAATTCAGGTAAAGA-A 23573
```

```
QY 301 GAAATAAAGAAATGAGAGATGAGAGATTTCTATGAAAAAGAAAGAGAAACATGTAGG 360
|||||
Db 23572 TAAATAAAGAAATGAGAGATGAGAGATTTCTATGAAAAAGAAAGAGAAACATGTGGG 23513
QY 361 TGAACAAAATAAGAGATATGATGATATATTTTATGAGAGGTGGTGAAGATTATTTTAGG 420
|||||
Db 23512 TGAACAAAATAAAGATATGATGATATATTTTATGAAAGGTGGTGAAGATTATTTTAGG 23453
QY 421 AGAGGAGAGAGAAATAGAAAAAGAAATGACATGGTGAATCTGAGAAAGATGAATTGTG 480
|||||
Db 23452 AGAGGAGAGAGAAATAGAAAAAGAAATGACATGGTGAATCTGAGAAAGATGAATTGTG 23393
QY 481 TTAAGATGAGAGAGAGAAAGAAAGAACTCCATGGCTAAAGTCTCGTAAAGAGATGAAAAAG 540
|||||
Db 23392 TTAAGAAATGAAGAGAGAAAGAAAGAACTCTATGACTAAAGTCTCGTAAAGAAACATGAAAAA 23333
QY 541 AAACAAAGAGAGAGAAAGAAAGAAAGCTAAATAGACTAACTATTTCGCAAAATTTTC 600
|||||
Db 23332 AAACAAAGAGAGAAATGAAGAAAGAAAGCTAAATAGACTAAATATTACCAAAATTTTC 23273
QY 601 TGTAGCCGACAAATACTATTGTGTCAGGTTTATTTGTGTTATTTTCTTTCAAGTCAAAAG 660
|||||
Db 23272 TGTAGCCGACAAATACTATTGTGTCAGGTTTATTTGTGTTATTTTCAAGTCAATGG 23213
QY 661 TTATTTCTTACATATCTCTAAATAATATAGCCGATACCAATTTTCCACACAT 713
|||||
Db 23212 TTATTTCTTACATATCTCTACATGTTTGTGTTAAATTTACTTATTTTAAATAT 23160
AC012394 94487 bp DNA linear PLN 12-OCT-2000
Arabidopsis thaliana chromosome 1 BAC F15M4 genomic sequence,
complete sequence.
AC012394
AC012394.3 GI:6554469
HTG.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE
1 (bases 1 to 94487)
AUTHORS
Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haase,B.J.,
Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
Barnstead,M.E., Bowman,C.L., White,O., Niernan,W.C. and Fraser,C.M.
Arabidopsis thaliana chromosome 1 BAC F15M4 genomic sequence
Unpublished
2 (bases 1 to 94487)
REFERENCE
2 (bases 1 to 94487)
AUTHORS
Lin,X. and Kaul,S.
TITLE
Direct Submission
JOURNAL
Submitted (27-OCT-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
3 (bases 1 to 94487)
REFERENCE
3 (bases 1 to 94487)
AUTHORS
Lin,X.
TITLE
Direct Submission
JOURNAL
Submitted (10-DEC-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
4 (bases 1 to 94487)
REFERENCE
4 (bases 1 to 94487)
AUTHORS
Town,C.D. and Kaul,S.
TITLE
Direct Submission
JOURNAL
Submitted (05-OCT-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
5 (bases 1 to 94487)
REFERENCE
5 (bases 1 to 94487)
AUTHORS
Town,C.D. and Kaul,S.
TITLE
Direct Submission
JOURNAL
Submitted (12-SEP-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
6 (bases 1 to 94487)
REFERENCE
6 (bases 1 to 94487)
AUTHORS
Town,C.D. and Kaul,S.
TITLE
Direct Submission
JOURNAL
Submitted (12-OCT-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
```

## COMMENT

On Dec 10, 1999 this sequence version replaced gi:6143858.  
Address all correspondence to: at@tigr.org

BAC clone F15M4 is from Arabidopsis thaliana chromosome 1  
The orientation of the sequence is from SP6 to T7 end of the BAC  
clone.  
Genes were identified by a combination of several methods: Gene  
prediction programs including Genscan+ (Chris Burge,  
<http://CCR-081.mic.edu/GENSCAN.html>), GeneMarkHM (Mark Borodovsky,  
<http://genemark.biology.gatech.edu/GeneMark/>), GlimmerA (a variant  
of GlimmerM, see Mihaela Pertea,  
<http://www.tigr.org/softlab/glimmerm.htm/glimmerm.html>, and  
GeneSplicer (Mihaela Pertea and Steven Salzberg, contact  
mpertea@tigr.org), searches of the complete sequence against a  
peptide database and the plant EST database at TIGR  
(<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to  
indicate the level of evidence for their annotation. Genes with  
similarity to other proteins are named after the database hits.  
Genes without significant peptide similarity but with EST  
similarity are named as unknown proteins. Genes without protein  
or EST similarity, that are predicted by more than two gene  
prediction programs over most of their length are annotated by  
hypothetical proteins. Genes encoding tRNAs are predicted by  
tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).  
Simple repeats are identified by repeatmasker (Arian Smit,  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

## FEATURES

## source

1. 94487  
/organism="Arabidopsis thaliana"

/mol\_type="genomic DNA"

/cultivar="Columbia"

/db\_xref="taxon:3702"

/chromosome="1"

/clone="F15M4"

/complement(967..1017)

/rpt\_family="AT rich"

/complement(1289..1338)

/rpt\_family="AT rich"

/complement(4053..4080)

/rpt\_family="AT rich"

/complement(6943..9510)

/gene="F15M4.1"

/note="identical to hydroxy methylglutaryl CoA reductase

(AA 1-592) [GB:X15032 [Arabidopsis thaliana]]

complement (join(<6943..7515,7607..7953,8115..8296,

8522..>9510))

/gene="F15M4.1"

complement (join(7255..7515,7607..7953,8115..8296,

8522..9510))

/gene="F15M4.1"

/codon\_start=1

/product="hydroxy methylglutaryl CoA reductase (AA 1-592) ;

9510-7255"

/protein\_id="AAF16652.1"

/db\_xref="GI:6554470"

/translation="MDLRRPKPPVPTNNNGSPRSYQRTSDDDHRRRATTIAPP  
PKASDALPLPLYNATVFFLFSVAYLLHRRDKIRYNTPLVVITITELGAIILI  
ASFIYLLGFGDFVQSIFSRASGDADLADTDDDDHRLVTCSPPTIVSVAKLPNP  
EPVLESPEDEBEIVKSVIDGVPISLESRLGDCCKRAISRREALQRTVASTIEGL  
PLDGFDPYSILGQCCMPGVYTIQIPVGIAGPLLDDGYEYSPMATTEGCLAVNRGRF  
KAMFISGGATSVLKDGMTRAPVPRFASARRASELKFPLENPFDTLAVVNRSRF  
ARLQSVKTIAGNATVRFCCSTGDAGNMVSKGVNVLEYLTDFFDMDDVIGISGN  
FCSDDKPAVNINIEGRKSVCEAVIRGEIVNKLKTSVAALVELNMLNKLASGAVG  
SLGGFNASINIVSFIATGQPAQNVESQCITMMEAINDKDHIHIVTWPFSIEVG  
TVGGGTQLASQACALLNLGVKGASTESPGNNARRLATI VAGAVLAGELSLMSAIAAGQ  
LVSRHMKVNRSSRDISGATTTTITTT"

complement (7258..7283)

/rpt\_family="(CAA)n"

complement(11154..11209)

/rpt\_family="AT rich"

complement(11551..11580)

/rpt\_family="AT rich"

complement(11933..12004)

## gene

/rpt\_family="AT rich"

12390..13268

/gene="F15M4.2"

/note="similar to unknown protein GB:AAD30603 [Arabidopsis

thaliana]"

join(<12390..12391,12491..12629,12730..12861,12929..13019,

13099..>13268)

/gene="F15M4.2"

join(12390..12391,12491..12629,12730..12861,12929..13019,

13099..13268)

/gene="F15M4.2"

/codon\_start=1

/product="unknown protein; 12390-13268"

/protein\_id="AAF16653.1"

/db\_xref="GI:6554471"

/translation="MFDVKALYVGKELWRETLQSGSRVYKQGLKSNWYEVKISY

PASIPARVSLQLLKNHEMLKINOMRLLNTEKLIFKAESLKEVKNKAGLNVLTLPE

EGIVAIPNSRRRYFIYINIVEEQLMGIPYSWSVLLVALCLVVSFIVPRSLPSLL

TKQQLRSRSHRGKDS"

14056..15506

/gene="F15M4.3"

/note="similar to cinnamoyl-CoA reductase GB:CAA56103

[Eucalyptus gunnii]"

join(<14056..14170,14277..14431,14516..14701,14819..14990,

15096..15285,15362..>15506)

/gene="F15M4.3"

join(14056..14170,14277..14431,14516..14701,14819..14990,

15096..15285,15362..15506)

/gene="F15M4.3"

/codon\_start=1

/product="putative cinnamoyl-CoA reductase; 14056-15506"

/protein\_id="AAF16654.1"

/db\_xref="GI:6554472"

/translation="NAVKQKVCVTGAGGFIASWLKFLLSRGYTVHGTVRDPCDEKND

HLRLDNASKNLKFKADLFDDEGLFSAIDGSGVFHIASVPFPEGEELIKPALTCYTK

NVLEACTETKQVKVYVSSIAAVYNPKWPQDVADEDCWSDTQVLSLEGWRYYL

AKTLFERALEKSNKFNADVTLCSFVIIGPRQLSTLSSSSGLLKPIKGGIKSLSD

ELYLDVRDVADALLLVYENREATGYICNSHLSYTLDSLMKELKNMKYPRNPFESFTE

VKEKEVRPLSAEKLKNLGNKFRPLEETIDDSVVSFEAAGDLPKA"

16281..16361

/rpt\_family="(CAT)n"

16955..18423

/gene="F15M4.4"

/note="contains Pfam profile: PF00076 RNA recognition

motif. [a.k.a. RRM, RBD, or RNP domain]"

join(<16955..17161,17385..17529,17639..17755,17844..17884,

17978..18215,18332..>18423)

/gene="F15M4.4"

join(16955..17161,17385..17529,17639..17755,17844..17884,

17978..18215,18332..18423)

/gene="F15M4.4"

/codon\_start=1

/product="putative RNA-binding protein; 16955-18423"

/protein\_id="AAF16655.1"

/db\_xref="GI:6554473"

/translation="MAYQVPVSGSGHYLNSPPGDTTFTTKVFGGLAWETQSETLRQHF

EYQGIILAVVIADNTKGRSGYGFVTRDPEARRACADPTPIIDGRANCNLASLG

RRPPLPVAVIPNMPDLLPRTLMCRVLEHLEIRISNHFHITTSKELSLIMVTAY

GPYMYQSQGLSPYMQOYLQVGVPCAVNSPVYQGLSQTIPNGHGYTAVQVGS

VPGSHILQLGFTVSTMTSSPALQAPYPSGIPGPAPVQSHIIVHSQFMQSTASDQ

TTT"

complement(19352..20843)

/gene="F15M4.5"

complement (join(<19352..19378,19455..19499,19581..19699,

19777..19852,19942..20024,20291..20356,20433..20536,

20615..20690,20777..>20843))

/gene="F15M4.5"

complement (join(19352..19378,19455..19499,19581..19699,

19777..19852,19942..20024,20291..20356,20433..20536,

20615..20690,20777..20843))

/gene="F15M4.5"

/codon\_start=1

/product="unknown protein; 20843-19352"

## repeat\_region

## gene

/rpt\_family="AT rich"

16955..18423

/gene="F15M4.4"

/note="contains Pfam profile: PF00076 RNA recognition

motif. [a.k.a. RRM, RBD, or RNP domain]"

join(<16955..17161,17385..17529,17639..17755,17844..17884,

17978..18215,18332..>18423)

/gene="F15M4.4"

join(16955..17161,17385..17529,17639..17755,17844..17884,

17978..18215,18332..18423)

/gene="F15M4.4"

/codon\_start=1

/product="putative RNA-binding protein; 16955-18423"

/protein\_id="AAF16655.1"

/db\_xref="GI:6554473"

/translation="MAYQVPVSGSGHYLNSPPGDTTFTTKVFGGLAWETQSETLRQHF

EYQGIILAVVIADNTKGRSGYGFVTRDPEARRACADPTPIIDGRANCNLASLG

RRPPLPVAVIPNMPDLLPRTLMCRVLEHLEIRISNHFHITTSKELSLIMVTAY

GPYMYQSQGLSPYMQOYLQVGVPCAVNSPVYQGLSQTIPNGHGYTAVQVGS

VPGSHILQLGFTVSTMTSSPALQAPYPSGIPGPAPVQSHIIVHSQFMQSTASDQ

TTT"

complement(19352..20843)

/gene="F15M4.5"

complement (join(<19352..19378,19455..19499,19581..19699,

19777..19852,19942..20024,20291..20356,20433..20536,

20615..20690,20777..>20843))

/gene="F15M4.5"

complement (join(19352..19378,19455..19499,19581..19699,

19777..19852,19942..20024,20291..20356,20433..20536,

20615..20690,20777..20843))

/gene="F15M4.5"

/codon\_start=1

/product="unknown protein; 20843-19352"

## gene

/rpt\_family="AT rich"

19352..20843

/gene="F15M4.5"

complement (join(<19352..19378,19455..19499,19581..19699,

19777..19852,19942..20024,20291..20356,20433..20536,

20615..20690,20777..>20843))

/gene="F15M4.5"

complement (join(19352..19378,19455..19499,19581..19699,

19777..19852,19942..20024,20291..20356,20433..20536,

20615..20690,20777..20843))

/gene="F15M4.5"

/codon\_start=1

/product="unknown protein; 20843-19352"

## gene

/rpt\_family="AT rich"

19352..20843

/gene="F15M4.5"

complement (join(<19352..19378,19455..19499,19581..19699,

19777..19852,19942..20024,20291..20356,20433..20536,

20615..20690,20777..>20843))

/gene="F15M4.5"

complement (join(19352..19378,19455..19499,19581..19699,

19777..19852,19942..20024,20291..20356,20433..20536,

20615..20690,20777..20843))

/gene="F15M4.5"

/codon\_start=1

/product="unknown protein; 20843-19352"

## gene

/rpt\_family="AT rich"

19352..20843

/gene="F15M4.5"

complement (join(<19352..19378,19455..19499,19581..19699,

19777..19852,19942..20024,20291..20356,20433..20536,

20615..20690,20777..>20843))

/gene="F15M4.5"

complement (join(19352..19378,19455..19499,19581..19699,

19777..19852,19942..20024,20291..20356,20433..20536,

20615..20690,20777..20843))

/gene="F15M4.5"

/codon\_start=1

/product="unknown protein; 20843-19352"

## gene

/rpt\_family="AT rich"

19352..20843

/gene="F15M4.5"

complement (join(<19352..19378,19455..19499,19581..19699,

19777..19852,19942..20024,20291..20356,20433..20536,

20615..20690,20777..>20843))

/gene="F15M4.5"

complement (join(19352..19378,19455..19499,19581..19699,

19777..19852,19942..20024,20291..20356,20433..20536,

20615..20690,20777..20843))

/gene="F15M4.5"

/codon\_start=1

/product="unknown protein; 20843-19352"

## gene

/rpt\_family="AT rich"

```
/protein_id="AAF1656.1"
/db_xref="GI:6554474"
/translation="MAAISPWSSPQSFNSPRVTITDSRRCSISAAISVLDSNBEQ
HRTSSRDHGWKMRDVMLOIASSVFLPLAISPAFAETNASEAFRVYTDITNKFEISI
POYSIATIGLGPDPTRNESGKVEAPFETLVSLGRDQWKPVGVTAKLIDSRASKGFI
YIETLONPEARKHLISALGMATNGWLNKLYTVIGQFTDEESAEOSSKIQKTVKSPF
FI"
repeat_region complement(20998..21026)
repeat_region rpt_family="AT_rich"
gene complement(21140..21167)
rpt_family="AT_rich"
22462..23090

Query Match 61.7%; Score 593.8; DB 8; Length 94487;
Best Local Similarity 94.7%; Pred. No. 6.4e-83;
Matches 641; Conservative 0; Mismatches 22; Indels 14; Gaps 2;

QY 1 TGTGTTCTCAGAAATAGCAGAAATATTATATAAAGCATGCAATCTCTTATAGATCG 60
Db |||||
QY 1830 TGTGTTCTCAGAAATAGCAGAAATATTATATAAAGCATGCAATCTCTTATAGATCG 1889
Db |||||

QY 61 CGAAGTTTAAAAAACATATAGATTGTTACAATATTACATGGCTTTTATTGGATAACA 120
Db |||||
QY 1890 CGAAGTTTAAAAAACATATAGATTGTTACAATATTACATGGCTTTTATTGGATAACA 1936
Db |||||

QY 121 TGACAAATATTATTATTTTCATGAGTTTATTATGGATAGCATGACAAATATTAAATAT 180
Db |||||
QY 1937 TGACAAATATTATTATTATTTTCATGAGTTTATTATGGATAGCATGACAAATATTAAATAT 1996
Db |||||

QY 181 CAGTGTTAATACATGTTTGTCTTAAATACATGCAATTTAAATCAGACATTTGTTT 240
Db |||||
QY 1997 TAGTGTTAATACATGTTTGTCTTAAATACATGCAATTTAAATCAGACATTTGTTT 2056
Db |||||

QY 241 TAAATCAAAATCTAATCTTATATCAACAGCATTCAGCGAAATTCAGGTAAAGA 300
Db |||||
QY 2057 TAAATCAAAATCTAATCTTATATCAACAGCATTCAGCGAAATTCAGGTAAAGA 2116
Db |||||

QY 301 GAAATTAAGAATGAGAGATAGAGAGATTTCTATGAAAAAGAAAGAGAGACATGTAGG 360
Db |||||
QY 2117 GAAATTAAGAATGAGAGATAGAGAGATTTCTATGAAAAAGAAAGAGAGACATGTGGG 2176
Db |||||

QY 361 TGAAC-AAATTAAGAGATATGATATATTTTATGAGAGGTGGTGAAGATTTATTAG 419
Db |||||
QY 2177 TGAACAAAAATAAGAGATATGATATATTTTATGAGAGGTGGTGAAGATTTATTAG 2236
Db |||||

QY 420 GAGAGGAGAGAGAAATAGAAAAGAAATGACATGCTGATCTGAAGAGATGAATTGT 479
Db |||||
QY 2237 GAGAGGAGAGAGAAATAGAAAAGAAATGACATGCTGATCTGAAGAGATGAATTGT 2296
Db |||||

QY 480 GTTAAAGATCAAGAGAGAGAAAGAACTCCATGGCTTAAAGTCTCTGTAAGAGATGAAAA 539
Db |||||
QY 2297 GTTAAAGATCAAGAGAGAGAGAAAGAACTCTATGGCTTAAAGTCTGTAAGAGATGAAAA 2356
Db |||||

QY 540 GAAACAAAAGAGAGAGAGAGAAAGAGAAAGGCTAAATAGACTAATCTTTGCCAAATTT 599
Db |||||
QY 2357 GAAATTAAGAAGAGAGAGAGAGAAAGAGAAAGGCTAAATAGACTAATCTTTGCCAAATTT 2416
Db |||||

QY 600 CTGTAGCCGACAAATCTATTTCGTCGAAGTTATTTTGTGTTCTTTTGAAGTCAAAA 659
Db |||||
QY 2417 CTGTAGCCGACAAAGTACTATTTCGTCGAAGTTATTTTGTGTTCTTTTCAACTCAAG 2476
Db |||||

QY 660 GTTATTCTTTACATATA 676
Db |||||
QY 2477 GTTATTCTTTACATATA 2493
Db |||||

RESULT 6
AC015450/c AC015450 100806 bp DNA linear PLN 15-JUN-2001
LOCUS Arabidopsis thaliana chromosome 1 BAC F14G6 genomic sequence,
DEFINITION complete sequence.
ACCESSION AC015450
VERSION AC015450.5 GI:12323968
```

HTG.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopses;  
REFERENCE 1 (bases 1 to 100806)  
AUTHORS Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J.,  
Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,  
Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.  
Arabidopsis thaliana chromosome 1 BAC F14G6 genomic sequence  
Unpublished  
TITLE 2 (bases 1 to 100806)  
JOURNAL Lin,X. and Kaul,S.  
REFERENCE Direct Submission  
AUTHORS Submitted (16-NOV-1999) The Institute for Genomic Research, 9712  
JOURNAL Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org  
TITLE 3 (bases 1 to 100806)  
AUTHORS Town,C.D. and Kaul,S.  
JOURNAL Direct Submission  
TITLE Submitted (19-JAN-2001) The Institute for Genomic Research, 9712  
JOURNAL Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org  
COMMENT On Jan 19, 2001 this sequence version replaced gi:12280761.  
Address all correspondence to:at@tigr.org

BAC clone F14G6 is from Arabidopsis thaliana chromosome 1  
The orientation of the sequence is from SP6 to T7 end of the BAC  
clone.  
Genes were identified by a combination of several methods: Gene  
prediction programs including Genscan+ (Chris Burge,  
http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHM (Mark Borodovsky,  
http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant  
of GlimmerM, see Mihaela Pertea,  
http://www.tigr.org/softlab/glimmerm.htm/glimmerm.html, and  
mpertea@tigr.org), searches of the complete sequence against a  
peptide database and the plant EST database at TIGR  
(http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to  
indicate the level of evidence for their annotation. Genes with  
similarity to other proteins are named after the database hits.  
Genes without significant peptide similarity but with EST  
similarity are named as unknown proteins. Genes without protein  
or EST similarity, that are predicted by more than two gene  
prediction programs over most of their length are annotated as  
hypothetical proteins. Genes encoding tRNAs are predicted by  
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).  
Simple repeats are identified by RepeatMasker (Arian Smit,  
http://ftp.genome.washington.edu/RM/RepeatMasker.html).

FEATURES  
Location/Qualifiers  
1..100806  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="1"  
/clone="F14G6"

misc\_feature 1  
/note="35000 nt before this point were not included in  
the submitted sequence, due to overlap with another BAC  
[F15M4]"  
5217..5774  
/gene="F14G6.1"  
/note="contains Pfam profile: PF00097 Zinc finger, C3HC4  
type (RING finger)"  
5217..5774  
/gene="F14G6.1"  
5217..5774  
/gene="F14G6.1"  
/codon\_start=1  
/product="putative RING zinc finger protein; S217-5774"  
/protein\_id="AAG51946.1"  
/db\_xref="GI:12323975"  
/translation="MARLLFRLLQEA NSTSPAE SPFN SLDVLILAVLLCA L TCIIG

```
gene
LIAVSRCAMLRRIASRRSDOTHPPVPAVAANKGLKKKVLRLSPKLTYSPPDSPPAEKLV
EACILCTEAAAGDELRLVLCQCGHFVHSCIDTWLGSSSHSCPSCRQILLVTRCHKCGGL
PGSSSGPPEPDRIKOREGDPNLP"
complement (8331. .10137)
/genes="F14G6.2"
/notes="N-term similar to N-term of NAM GB:CAA63101
[petunia x hybrida] (apical meristem formation), CUC2
GB:BAAL9529 [Arabidopsis thaliana], GRAB2 protein
GB:CAA09372 [Triticum sp.]"
complement (join(<8331. .8855,9526. .9800,9933. .>10137))
/genes="F14G6.2"
complement (join(8331. .8855,9526. .9800,9933. .10137))
/genes="F14G6.2"
/codon_start=1
/product="unknown protein; 10137-8331"
/protein_id="AAG51953.1"
/db_xref="GI:12323982"
/translation="MMLAVEDVLSLAGEERNELGPPGFRFPTDEELITFYLASKI
FHGGLSGIHSVDLNRCEPMEIPENAKMGEREMFYSLDRKYPTGLTRNATAGY
WKATGDXKEVFGGGQLVGMKTLVYKGRAPRGLTKWVHEVRLDNDHSHRHTCK
EWMVLCRVNKTGDRKNVGLIHQI18LHNHLSLSTHHHHEALPULIIPSNNKLTNF
PSLLYDDPHQNTNNNNFLHSGSHNIDELKALINPVVSQLGIIPSGNNNDEDDFD
FNLVYQTEQSSNGNEIDVRDYLENPLFQEASVGLLGFSSSPGPLHMLDSCPGLGQL"
complement (15417. .18176)
/genes="F14G6.3"
/notes="similar to phosphate transporter GB:BAAL20522
[Catharanthus roseus]; contains Pfam profile: PF00083
Sugar (and other) transporter"
complement (join(<15417. .16305,17467. .>18176))
/genes="F14G6.3"
complement (join(15417. .16305,17467. .18176))
/genes="F14G6.3"
/codon_start=1
/product="putative phosphate transporter; 18176-15417"
/protein_id="AAG51941.1"
/db_xref="GI:12323970"
/translation="MPELSLSALDAARTQWHFKAIIVAGMGLFTDADYDLFCIAPIM
KMSQIYHKDSIGTALLSYAIALGLTALGQIFGYLGDVRGRKRYVGLSLLIIVF
SPFGCFYSCTTRSCVMVSLGFLFVGLGIGDGYPLSATIMSEFANKRTRGAFTAA
VFSMQGLGILWMSAVTMVCLAPKNAGEGSEKTNVAGLETAPLAPESDIARWILMIG
ALPAAULTFYVLMMPETARYTALVENNVQAAKMORVMSVSMISOITEDSSSELBP
PSSSVKLFSPRLSLHGRDLFAASANWELVDVVFYTSNLLLSQIFENFSKNPLSNVY
YDSAPVAKLAIAVACSTIPGWFTVYFIDIKLRVKIOMWGFFLMAVYLVAGIPIYS
WWSRKEKNTGMVLYGLIFPFSNFGPNTTPIIPAEFLPARKFRSTCHGISGAAGK
GAIVGVFLMATRHHEDGPDVQRVRIAFILILGVCIAGMIVTYLFTRETMRSLR
ENEDIVTSAGSSPANELLRRQY"
complement (18673. .19301)
/genes="F14G6.4"
/notes="predicted by genscan+"
complement (join(<18673. .18926,19124. .>19301))
/genes="F14G6.4"
complement (join(18673. .18926,19124. .19301))
/genes="F14G6.4"
/codon_start=1
/product="hypothetical protein; 19301-18673"
/protein_id="AAG51943.1"
/db_xref="GI:12323972"
/translation="MNAENNQTTHSKVISHVFCTGTAKLGSVGPPIGLVDIGVSEV
AVIFRVLPGIEKNQKICEIQREGRVCIQVIPLEIAPSDTGCLYRQVQVQLCPPG
PFSITFNLQVDPDRLPFSNFRSDGIFEVVVVVKLGVRIPTS"
20920. .22411
/genes="F14G6.5"
join(<20920. .20986,21073. .21148,21227. .21330,21407. .21472,
21739. .21821,21911. .21986,22064. .22182,22264. .22308,
22385. .>22411)
/genes="F14G6.5"
join(20920. .20986,21073. .21148,21227. .21330,21407. .21472,
21739. .21821,21911. .21986,22064. .22182,22264. .22308,
22385. .22411)
/genes="F14G6.5"
/codon_start=1
/product="unknown protein; 20920-22411"
```

```
gene
/protein_id="AAG51945.1"
/db_xref="GI:12323974"
/translation="MAAISPLUSSQSPSNRPVTTIDRRCSISIAAISVLDSNBEQ
HRISRDHVMKRRDVMQIASSVFLPLAISPAFAETNASEAFRYVTTETNKFETSI
VQSTAITGLGPDFTRMESFGKVEAFETLVSGLDRSQKQVPGVTAKLIDSRASKGFY
PYEYTLQNPGEARKHLYSAGMATNGWYRLYTVTQGTDEESAEQSSKIQTVKFSR
FI"
complement (23340. .24808)
/genes="F14G6.6"
/notes="contains Pfam profile: PF00076 RNA recognition
motif. (a.k.a. RRM, RBD, or RNP domain)"
complement (join(<23340. .23431,23548. .23785,23879. .23919,
24008. .24124,24234. .24378,24602. .>24808))
/genes="F14G6.6"
complement (join(23340. .23431,23548. .23785,23879. .23919,
24008. .24124,24234. .24378,24602. .24808))
/genes="F14G6.6"
/codon_start=1
/product="putative RNA-binding protein; 24808-23340"
/protein_id="AAG51948.1"
/db_xref="GI:12323977"
/translation="MAYQVPVGSGFHYLNSPFGDTTFTKFVGLAWETQSETLRQH
EYGGTILEAVIADKNTGRSGKGYFVTRDPAARRACADPTPIIDGRANCNLASIG
RRPPLPVAVINPMDLLPRTLEMCRVLEVHLEIIRISNHFHTTISKELSIIMVTAY
GPEYMSQSQGLYSPVMOQYLVQVPGAVNSPVYQGLSQTIPNGHGYTAVQGS
VFGSHILQGGFTVSTMTSSMPALQAPPSGIPGPVPVQSHIIVHSPQFMQSTASDQ
TTR"
complement (26257. .27736)
/genes="F14G6.7"
/notes="similar to cinnamoyl-CoA reductase GB:CAA56103
[Eucalyptus gunnii]; contains non-consensus GG acceptor
splice site at exon 4"
complement (join(<26257. .26401,26478. .26667,26773. .26935,
27062. .27247,27332. .27486,27593. .>27736))
/genes="F14G6.7"
/notes="contains nonconsensus splice site"
complement (join(26257. .26401,26478. .26667,26773. .26935,
27062. .27247,27332. .27486,27593. .27707))
/genes="F14G6.7"
/notes="contains nonconsensus splice site"
/codon_start=1
/product="putative cinnamoyl-CoA reductase; 27707-26257"
/protein_id="AAG51951.1"
/db_xref="GI:12323980"
/translation="MAVKOKVCVTGAGGFIASWLKFLSRGYTHGVTRVDRPCDEKND
HUKLDNASKNLKLFKADLFDDEGLFSADIGCGVPHIASVPVPEGEELIKPALTGTK
NLVEACTETKVQKVVVSSIAAVVNPWKVPQVAKDECDWSQTYLHSLRYVYLAAT
LTEREALWSKRNFDVVTLCPSVIIGPRLQSLNSSSLGLKFKIKGKISLSDLY
LVDVRDADALLLVENREATGRYICNSHSLYTDLSLMEKLNKMPKRNPFESFTEVKE"
Query Match 61.7%; Score 593.8; DB 8; Length 100806;
Best Local Similarity 94.7%; Pred. No. 6.3e-83;
Matches 641; Conservative 0; Mismatches 22; Indels 14; Gaps 2;
QY 1 TGTGTTTCTCAGAAATAGCAGCAAAATATTTATAAAGCATGCAATCTCTTATAGATCG 60
Db 39933 TGTGTTTTTCAGAAATAGCAGCAAAATATTTATAAAGCATGCAATCTCTTATAGATTG 39874
QY 61 CGAAGTTTAAAAAACAATAGAAATGTTTACATAATTACATGGGTTTTTTATTCGATAACA 120
Db 39873 CGAAATTTAAAAA -----TTACAATATTACATGGGTTTTTTATTCGATAACA 39827
QY 121 TGCAAAATATTTATTTATTTTCATGAGTTTTTTATTTGGATGACATGACAAATATTATAT 180
Db 39826 TGCAAAATATTTATTTATTTTCATGAGTTTTTTGTTGGATGACATGACAAATATTATAT 39767
QY 181 CAGTGTTAATACATGTTTTTGTCTTAAATAACATGCATTTTAAATATCAGACATTTGTTT 240
Db 39766 TAGTGTTAATACATGTTTGTCTTAAATAACATGCATTTTAAATATCAGATTTTGTGT 39707
QY 241 TAAATCAAATCTAAATCTCTTATATACAAACGACATTCAGCGGAAATTCAGGTAAAAAGA 300
Db 39706 TAAATCAAATTTAAATCTCTTATATACAAACGACATTCAGCGGAAATTCAGGTAAAAAGA 39647
```



```
exon      26786..26947
           /gene="AT4g08710"
           /number=3
intron    26948..27033
           /gene="AT4g08710"
           /number=3
exon      27034..27477
           /gene="AT4g08710"
           /number=4
intron    27478..27518
           /gene="AT4g08710"
           /number=4
exon      27519..27800
           /gene="AT4g08710"
           /number=5
gene      28303..31953
           /gene="AT4g08720"
gene      complement(join(28303..28494,28565..28761,29626..30162,
30333..31953))
           /gene="AT4g08720"
CDS       complement(join(28303..28494,28565..28761,29626..30162,
30333..31953))
           /gene="AT4g08720"
           /note="similarity to hypothetical protein F26B6.15 -
Arabidopsis thaliana, PID:g3242713"
           /codon_start=1
           /product="putative protein"
           /protein_id="CAB82108.1"
           /db_xref="GI:7321061"
           /db_xref="UniProt/TREMBL:Q9LDY3"
           /translation="WVEKVLIYFRKDRAYSLMMNGKITLSMLESICSLGLDESKV
KLOKNYNAVMLGLTQSIHSDDEDVGVYISSGEMWSRYPLTVDVDPPELPELSKVE
VYNKSIGKNYDGLSGNEDHEDMNAITIVIGELDQGTLELNGAIVVDDTDDEHIGSRV
DRDDDRGGDEYVPEPPDVESSQPKKEWDGIGLTROEPFSPRAALQEVVDYKGFANSF
DYVIKCDKERYLVITCSKANCOWRIQASCVCQDTSIYSIRRYNKHSCSTRISKSSTRMF
KRKGTPELVAALHPTDFPGLLETPPKVIMELVOTKLGVKVSYSTALRGKROAVCDLK
GAEDSYKDINCVLNMLKKVNDGTVTYMKLDSGKFQYLFIALGASIEGFQAMRKVII
VQATHLKNYGGVGLVPASARDPNRHHYIYAVGLDGENDASGWFEKLLSVVPTPE
LVFMSDRNSLLIGIRNAYTAAHHGYCVHLSQNVKAHSTNINRDVLAWRFMELSRIV
TWSEFEYRVRVFRVRYPTAVNYLEDSTVKEKWSVDNKFVPLVENVYLNHLWDAEKLIY
VTELNTFRLEYNVRVGEVGEYLVNLFHTCKSCGVFDIOKYPCTIHALAAIFVDSFTTR
TRGMELHELVSYKYWAEALWALAYGTIYLVPGRSOWDI PDDIKALKVLP LPRKKKGR
TKVLRTPSTGERPKRQORTONKRRPRQSLQWLLFGTRTWDRLCESDPWYGEWMSAKRI
MQOLEVAMGEIPPLPCPGGRILDIISEKDGKGRFYECTNRYKNDGLHRIKLWDKA
MVEEVNRLREQVDNHQKIQSUEYINQQMRSLKDEIQKXNGTLWLRICCGE"
           /genes="AT4g08720",
           complement(28303..28494)
           /number=1
intron    complement(28495..28564)
           /gene="AT4g08720"
exon      complement(28565..28761)
           /gene="AT4g08720"
           /number=2
intron    complement(28762..29625)
           /gene="AT4g08720"
           /number=2
exon      complement(29626..30162)
           /gene="AT4g08720"
           /number=3
intron    complement(30163..30332)
           /gene="AT4g08720"
           /number=3
exon      complement(30333..31953)
           /gene="AT4g08720"
           /number=4
gene      34618..36007
           /gene="AT4g08730"
gene      complement(join(34618..34852,35539..35690,35828..36007))
           /gene="AT4g08730"
CDS       complement(join(34618..34852,35539..35690,35828..36007))
           /gene="AT4g08730"
           /codon_start=1
```

```

/product="hypothetical protein"
/protein_id="CAB82109.1"
/db_xref="GI:7321062"
/db_xref="UniProt/TREMBL:Q9LDC3"
/translation="MEGFTNFIQVMEKTKVDLLARKVLDLTSKMDLCATNVDDLTSKWN
LCATNSDLKAMLTSMKGNVPAPKTLMAAASDSESDSSDDIELCFRFLIKIVIO
LRPSYFIILIPAYIDLKEGANKALELNGSDMAGKELVVKLTALLMRDSYLGYSNCGGIGG
RFGRPGSVGGIFGGRCGGLGQYAAQV"
complement(34618..34852)
/gene="AT4g08730"
/number=1
complement(34853..35538)
/gene="AT4g08730"

Query Match      60.2%; Score 579.2; DB 8; Length 104386;
Best Local Similarity 90.9%; Pred. No. 1.1e-80;
Matches 651; Conservative 0; Mismatches 58; Indels 7; Gaps 3;

QY      1  TGTGTTTCTCAGAAAATAGCACGAAATATTATATAAAAGCATGCAATTCCTTTATAGATCG 60
Db      87837 TGTGTTTCTCTGAAATAGCACCAGATATATTATAAAAGCATGCAATTCCTTTATAATCG 87778

QY      61  CGAAGT-TTAAAAAACATATAGAAATTGTTACAAATATTACATGGGTTTTTATTGGGATAAC 119
Db      87777 CGAAGTGTTTAAGAAACATATAGAAATTGTTACAAATTTACATCGGTTTTTATTGGATAAC 87718

QY      120 ATGACAAATATTATTATTATTTTCATGAGTTTTTATTGGATAGCATGACAAATATTAATATA 179
Db      87717 ATGACTAATATTATATATATTTTCATGAGTTTTTGTTCGATAGCATGACAAATATTAATATA 87658

QY      180 TCAGTGTTAATACATGTTTTTGTTCCTTAAAAATACATGCATTTTAAAAATCAGACATTTGTT 239
Db      87657 TCAGTGTTAATAACATGTTTTTGTTCCTTAAAAATACATGCATTTTAAAAATCAGACATTTGTT 87598

QY      240 TTAATAATCAAACTTAATCTCTTATATACAAACGACATTCACGCGAAAAATTCAGTAAAAAAG 299
Db      87597 TTAACATCAAACTTAATCTCTTATATACAAACGACATTCATGGAGAAATTCAGCGCAAAAAG 87538

QY      300 AGAAAAATAAGAAATGAGAGATAGAGAGATTTCTATCGAAAAAGAAAAGAGAGAAACATGTAG 359
Db      87537 AGAAAAATAAGAAATGAGAGATAGAGAGATTTCTATCGAAAAATAAGAGAGGAAACATGTGG 87478

QY      360 GTGACAAAAATAAGAGATATGATGATATATTTTATGAGAGGGTGGTGAAGATTTATT--- 416
Db      87477 GTGAACAAAAATAAGAGATATGATGATATATTTTATGAGAGGGTGGTGAAGATTTATTAGG 87418

QY      417 --TAGGAGAGGGAGAGAGAAATAGAAAAAGAAAATGACATGGTGAATCTGAAGAAAGATGA 474
Db      87417 AGAGGGAGAGAGAAATAGAAAATAGAAAAGAAAATGACATGGTGAATCTGAAGAAAAGATGA 87358

QY      475 ATTGTGTTAAAGATGAAGAGAGAGAAAGAACTCCATGGCTTAAAGTCTCGTAAAGAGAGATG 534
Db      87357 ATTGTGTTAAAGATGAAGAGAGAGAGAAAGAACTCCATAGCTTAAAGTCTCGTAAAGAACATG 87298

QY      535 AAAAAAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 594
Db      87297 AAAAAAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 87239

QY      595 AATTTCTGTAGCGGCAAAATACTATTGTGTCAGAGGTATTTTGTGTTATTTCTTTTCAAGT 654
Db      87238 AATTTCTGTAGTACACAGTACTATTGTTCCNAGGTATTTGTTTATTCTTTTCAAGT 87179

QY      655 CAAAAGTTATTTCTTACATATATCTCTTAAAAATATAGCCGATACCAATTTTTCACA 710
Db      87178 CAAAGGTTATTTGTATATATCTCTTTTATAAAAAAAAATAATAACGTTTGCAA 87123
```

```
RESULT 8
ATCHRIV25/c
LOCUS      ATCHRIV25      179771 bp      DNA      linear      PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 25.
ACCESSION  AL161513
VERSION     AL161513.2  GI:7267524
```

## KEYWORDS

## SOURCE

## ORGANISM

Arabidopsis thaliana (thale cress)  
 Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi  
 1 (bases 1 to 26363; 134798 to 179771)  
 Spiegel, L.A., Huang, E.N., Nascimento, L.U., de la Bastide, M.,  
 Vil, D.M., Preston, R.R., Matero, A., Shah, R., O'Shaughnessy, A.,  
 Rodriguez, M., Shekher, M., Schutz, K., See, L.H., Swaby, I.,  
 Habermann, K., Dedhia, N.N., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.  
 Unpublished

## JOURNAL

## REFERENCE

## AUTHORS

2 (bases 25642 to 34296)  
 Robben, J., Grymonprez, B., Bastiaens, I., Volckaert, G., Mewes, H.W.,  
 Lemcke, K. and Mayer, K.F.X.

## JOURNAL

## REFERENCE

## AUTHORS

3 (bases 33706 to 33707)  
 Lamar, B., Stoneking, I., Stumpf, J., Mewes, H.W., Lemcke, K. and  
 Mayer, K.F.X.

## JOURNAL

## REFERENCE

## AUTHORS

4 (bases 1 to 179771)  
 EU Arabidopsis sequencing project.

## JOURNAL

## REFERENCE

## AUTHORS

Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer  
 Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:  
 lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project  
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge  
 Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,  
 E-mail: michael.bevan@bbsrc.ac.uk

## COMMENT

Information on performance of analysis and a more detailed  
 annotation of this entry and other sequences of chromosomes 3, 4  
 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>  
 this fragment has an overlap with ATCHRIV24 at the 5' end and an  
 overlap with ATCHRIV26 at the 3' end.

## FEATURES

## source

1..179771  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /variety="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="4"  
 8902..10347  
 /gene="AT4g08820"  
 join(8902..9093,9193..9553,9686..10347)  
 /gene="AT4g08820"  
 /note="similarity to various hypothetical proteins of  
 Arabidopsis thaliana Chromosome II, III and IV"

## gene

## CDS

/codon\_start=1  
 /product="putative protein"  
 /protein\_id="CAB78007.1"  
 /db\_xref="GI:7267525"  
 /db\_xref="Uniprot/TREMBL:Q9LE84"  
 /translation="MSVSNKGSVGGGLKPEENVIDDFAFVRVRVLEFPNGDEGPVIT  
 IGEVLEAMNGLWKRCMIVRFIRFELEEYLAALTGGPMRVFGNYLVMQDSNPF  
 PLRDIIVTVTVRLTNIPVNYHRCLEELIARGLKLKLDLNTITFGRGRFARVIC  
 EVNLAKPLKTVLINDRIFYVAVEVEEGFTVRQSGIRVEKPAQKVMFAAGSGGRS  
 NRSRLRELPKNGVEISNRFCGLBEDMYSAEIGVAILEGSKENKYGKGNRKESGVT  
 QVKEAQNVDTFKKGKNGKGPWKNGGLIIMEPNGPRQKQSVNKPVRGLIFGPVK  
 GESSNLVDALSANGKRLRVEQRSVGRPGGAFVSGLETSGEMDMSELPQNNNQIQQ  
 SKELLCDDTAIVV"  
 8902..9093  
 /gene="AT4g08820"  
 /number=1  
 9094..9192  
 /gene="AT4g08820"  
 /number=1  
 9193..9553  
 /gene="AT4g08820"  
 /number=2  
 9554..9685  
 /gene="AT4g08820"  
 /number=2  
 9686..10347

## exon

## intron

## exon

## intron

## exon

/gene="AT4g08820"  
 /number=3  
 10891..114348  
 /gene="AT4g08830"  
 join(10891..11025,11244..11393,11430..12120,12361..13089,  
 13162..13585,13634..14348)  
 /gene="AT4g08830"  
 /note="similarity to various putative reverse  
 transcriptases of Arabidopsis thaliana  
 Contains Aminoacyl-transfer RNA synthetases class-II  
 signatures AA699-708"  
 /codon\_start=1  
 /product="putative protein"  
 /protein\_id="CAB78008.1"  
 /db\_xref="GI:7267526"  
 /db\_xref="GOA:Q9LDL8"  
 /db\_xref="Uniprot/TREMBL:Q9LDL8"  
 /translation="MSPDLSLEFGWENRVSILDMGFSGNKYTKWGRVENVTFVAKRLD  
 RLKKNRDVFQDVKKKEELMREIQVQDADLVHQSDMLRKEEDLIKAFDKSREKWI  
 ALGDRNTNYFHTTIIIRRNQRVERLHNEGTWIFDAKELEALAVQYQRLYSMDVL  
 OVVNPLPDGLERLRTREEVLSLNKPLATEVAIRSMGKVKAPGPGYQPIFYQSCW  
 EYVGNVIRFALDPFSSGILPQOTNDALVVLIPKPKPERMNOFRPISLCNAIPKMIT  
 KMVLRLKVIKGLIGPSQSSFIPGRSLDNIIVVQRAVHSLNRKKGRKRLCHMIDR  
 AVAKEMKSGISLQGGPKISHICFADDLILFAEASVQIRVIRRILETFCIASGQKVS  
 LDKSKIFFKSNVSRDLKLTISKESGISTRELKYLGMPTLQRIINRDKTDFEVLRSV  
 SRLAGWKGRSLSFAGRLTTLTKVSLIPIHTMTSTISLPQSTLEGLDKLARVFLIGSSA  
 EKKKLHLVADWRVCLPKSEGGIGIRTSCMKNKALSVKQWRLINDRYSLKARILRSKY  
 RVGLREVSRGSRWVNGRDILFWSDNWLSHEALINRAVIEIPNSEKEKLRVKDLWAN  
 GHDKLDLKEPIYSIHTRLUELAADVDSVTDGARDLSWGSADGVFTVKSAYRLITSD  
 LGPRNNAAFFDLRLRVVALERVKTLFHHIGDTSVCQCKGGDETILHLVKDCPSIAG  
 IWRRLVQVQVSRYDFNGSLFGMLVNLGKNKAETGYAWATLFAIVVMWSKRCGYVF  
 GEYKCRDRVKFFRDAAEVSHAHAIHSQNGRLTRVERLVANWKPDPCEWVKLNTDGA  
 SRNLGLATTGGVLRDGIHWCQGFALDQVCSAPLAELGWVYGLYTMAMERRFRIVE  
 LEVDSLVVGFLLTTGISDTHSLSLVRLCHGFL"  
 10891..11025  
 /gene="AT4g08830"  
 /number=1  
 11026..11243  
 /gene="AT4g08830"  
 /number=1  
 11244..11393  
 /gene="AT4g08830"  
 /number=2  
 11394..11429  
 /gene="AT4g08830"  
 /number=2  
 11430..12120  
 /gene="AT4g08830"  
 /number=3  
 12121..12360  
 /gene="AT4g08830"  
 /number=3  
 12361..13089  
 /gene="AT4g08830"  
 /number=4  
 13090..13161  
 /gene="AT4g08830"  
 /number=4  
 13162..13585  
 /gene="AT4g08830"  
 /number=5  
 13586..13633  
 /gene="AT4g08830"  
 /number=5  
 13634..14348  
 /gene="AT4g08830"  
 /number=6  
 16002..16053  
 18436..20750  
 /gene="At4g08840"  
 join(18436..19424,19870..20032,20124..20287,20492..20750)  
 /gene="AT4g08840"  
 /note="similarity to probable regulatory protein YGL014w"

## repeat\_region

## gene

## CDS



-Saccharomyces cerevisiae, PIR2:S64016  
Contains Prokaryotic membrane lipoprotein lipid attachment  
site AA82-92"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB78009.1"  
/db\_xref="GI:7267527"  
/db\_xref="GOA:Q9LDW3"  
/db\_xref="UniProt/trEMBL:Q9LDW3"  
/translation="MDQFFPGDLRQGSFTDLGFGFPLTSVSNQHFSGDRTNP  
FLNRLKDTTSLMADGDMGLCNLSKMSISDRSNFFNHSPSGYCYQGRSSPHG  
EASSMRGFGVGDYHRFDQDLRVASPHGESAMSSVYGDGYHRLRLFLAQEASNP  
NRCFTENSLNLDYMLEHFNQOIREDPSLVPOKSPLAHEERILPFSAMGSR  
ELDGSACMKNKEDSLDLSMVDSSYGYLMMAKQDGLCKLQKFVDEGNFVDMIFK  
EVINVEIGTDFPGNYLQKLEVCNEEQRTQILRLTSLKPGLLVKIISINNYGRV  
QKLEIATTKQISLVKLSALVPGLFLFRELNGNHVILNCLKFFSPNDKNFLEAATK  
FCIEIATTHGCCVLQRCVSYSGEHEKLVDIEISRNLSLLAQDPGNYLVQVIEKK  
VGVNVLFELRGNYVKLATQKFGSHVVEKCLRYYPESRSQIVNELVSLNFGVLLQDP  
YANYVIQCALSRTKVS"  
18436..19424  
/gene="At4g08840"  
/number=1  
19425..19869  
/gene="At4g08840"  
/number=1  
19870..20032  
/gene="At4g08840"  
/number=2  
20033..20123  
/gene="At4g08840"  
/number=2  
20124..20287  
/gene="At4g08840"  
/number=3  
20288..20491  
/gene="At4g08840"  
/number=3  
20492..20750  
/gene="At4g08840"  
/number=4  
20751..20750  
/gene="At4g08850"  
complement(join(23831..24120,24787..27580))  
/gene="At4g08850"  
complement(join(23831..24120,24787..27580))  
/gene="At4g08850"  
/notes="Est GB:F14413 is reverse and marks the end of the  
gene  
similarity to receptor protein kinase-like protein -  
Arabidopsis thaliana, PIR2:T05898  
Query Match 60.2%; Score 579.2; DB 8; Length 179771;  
Best Local Similarity 90.9%; Pred. No. 9.8e-81;  
Matches 661; Conservative 0; Mismatches 58; Indels 7; Gaps 3;  
QY 1 TGTGTTTCTCAGAAATAGCAGCAATATTTATATAAAGCATGCAATCTCTTATAGATCG 60  
|||||  
DB 17747 TGTGTTTCTCAGAAATAGCAGCAATATTTATATAAAGCATGCAATCTCTTATAATCG 17688  
QY 61 CGAAGT-TTAAAAAACAATAGAAATGTTTACAAATATTACATGGGTTTTTATTGGATAAC 119  
DB 17687 CGAAGTGTTTAAGAAACATAGAAATGTTTACAAATATTACATGGGTTTTTATTGGATAAC 17628  
QY 120 ATGACAAATATTATTATTTCATGAGTTTTTTATTGGATAGCATGACAAATATTATAATA 179  
DB 17627 ATGACTAATATTATATATTTTCATGAGTTTTTGTGGATAGCATGACAAATATTATAATA 17568  
QY 180 TCAGTGTTAATACATGTTTGTCTTAAATACATGCAATTTTAAATCAGACATTTGTT 239  
DB 17567 TCAGTGTTAATACATGTTTGTCTTAAATACATGCAATTTTAAATCAGACATTTGTT 17508  
QY 240 TTAATAATCAAACTTAATCTCTTATATACAAACGACATTTGACGGAATTCAGGTAAAAAG 299  
|||||

DB 17507 TTAACATCAAACTTAATCTCTTATATACAAACGACATTTGAGGAATTCAGGCAAAAAG 17448  
QY 300 AGAAATAAAGAAATGAGAGATAGAGATTTCTATGGAATAAAGAGAGAACATGTAG 359  
|||||  
DB 17447 AGAAACAAGAAATGAGAGATAGAGATTTCTATGGAATAAAGAGAGAACATGTGG 17388  
QY 360 GTGACAAATAAAGAGATATGATATATTTTATGAGAGGTGGTGAAGATTATTT--- 416  
DB 17387 GTGAACAAATAAAGAGATATGATATATTTTATGAGAGGTGGTGAAGATTATTTAGG 17328  
QY 417 --TAGGAGAGCGGAGAGAGAAATAGAAAAAGAAAAATGACATGCTGAATCTGAAGAACATGA 474  
DB 17327 AGAGGAGAGAGAAATAGAAATAGAAAAAGAAAAATGACATGCTGAATCTGAAGAACATGA 17268  
QY 475 ATTGTTTAAAGATGAAGAGAGAAAGAACTCCATGGCTAAAGTCTCGTAAAGAAAGATG 534  
DB 17267 ATTGTTTAAAGATGAAGAGAGAAAGAACTCCATAGCTAAAGTCTCGTAAAGAAACATG 17208  
QY 535 AAAAGAAACAAGAGAGAGAAAGAGAAAGAGAAAGGCTAAAAATAGCTAACTATTGCCAA 594  
DB 17207 AAAAATAAATAAGAGAGAAAGAGAAAGAGAAAGGCTAAAAATAGACTAA-TATTGCCAA 17149  
QY 595 AATTCTGTAGCGCAAAATACTATTGTGTCGAAGTTATTGTTGTGTTATTTCTTTTCAAGT 654  
DB 17148 AATTCTGTAGTAGACAGTACTATTGTTCCAGGTTATTGTTTATTTCTTTTCAAGT 17089  
QY 655 CAAAAGTTAATTTCTTACATATATCTCTAAAAATATAGCCGATACCAATTTTCCACA 710  
DB 17088 CAAAGTTATTGTTATATATATCTCTTTTATAAAAAAATAATAACGTTTGCAA 17033  
RESULT 9  
AF071527 95519 bp DNA linear PLN 06-MAY-1999  
LOCUS Arabidopsis thaliana BAC F9H3, from chromosome IV near 18.8 cm,  
DEFINITION complete sequence.  
ACCESSION AF071527  
VERSION AF071527.1 GI:3236479  
KEYWORDS HTG.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
REFERENCE 1 (bases 1 to 95519)  
AUTHORS Huang, E.N., Parnell, L.D., de la Bastide, M., Schutz, K.,  
Habermann, K., Dedhia, N.N. and McCombie, W.R.  
Genomic sequence of Arabidopsis thaliana BAC F9H3, chromosome IV,  
18.8 cm  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 95519)  
AUTHORS Huang, E.N., Parnell, L.D., de la Bastide, M., Schutz, K.,  
Habermann, K., Dedhia, N.N. and McCombie, W.R.  
Direct Submission  
JOURNAL Submitted (10-JUN-1998) Cold Spring Harbor Laboratories, Lita  
Annenberg Hazen Genome Center, 1, Bungtown Road, Cold Spring  
Harbor, NY 11724, USA  
REFERENCE 3 (bases 1 to 95519)  
AUTHORS Parnell, L.D. and McCombie, W.R.  
Direct Submission  
JOURNAL Submitted (31-DEC-1998) Lita Annenberg Hazen Genome Sequencing  
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring  
Harbor, NY 11724, USA  
REMARK Arabidopsis thaliana BAC F9H3 from chromosome 4, near 18.8 cm  
REFERENCE 4 (bases 1 to 95519)  
AUTHORS Parnell, L.D. and McCombie, W.R.  
Direct Submission  
JOURNAL Submitted (01-FEB-1999) Lita Annenberg Hazen Genome Sequencing  
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring  
Harbor, NY 11724, USA  
REMARK Arabidopsis thaliana BAC F9H3 from chromosome IV near 18.5 cm  
REFERENCE 5 (bases 1 to 95519)  
AUTHORS Parnell, L.D.

**TITLE** Direct Submission  
**JOURNAL** Submitted (06-MAY-1999) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724

**REMARK** Arabidopsis thaliana BAC F9H3 from chromosome IV near 18.8 cM

**COMMENT** BAC F9H3 is assigned between YACs C1C47 and C1C8B1 and maps to near 18.8 cM on the Lister & Dean RI map. Position 1 of F9H3 is oriented toward the centromere and position 95519 is oriented toward the telomere. For more information on the mapping, sequencing and annotation of F9H3, please see <http://www.cshl.org/Arabidopsis/F9H3-titlepage.html>. A graphic view of our annotation is also available at this url. Gene models are built with exons predicted by GenScan (<http://ccr-081.mit.edu/genefinder>) and MZEF (<http://www.cshl.org/genefinder>) and GRAIL (<http://compbio.ornl.gov/tools/index.shtml>) and with splice sites predicted by NetPlantGene (<http://www.cbs.dtu.dk/netgene/cbsnetgene.html>). Alternate exons not used in building the gene models are presented on the web pages associated with F9H3. Genes are numbered according to the scheme BAC.gene number. Typically, these numbers progress from 1 upwards as one moves from position 1 of the BAC. Protein sequences encoded by the genes are assigned to a functional category with the aid of similarity searches and comparison to the Prosite (<http://expasy.hcuge.ch/sprot/prosite.html>) and Pfam (<http://pfam.wustl.edu/>) libraries. A description of these categories can be found at <http://muntjac.mips.biochem.mpg.de/Arabi/>. Genomic repeats are typically located by TBLASTX analysis and an attempt is made to classify the function of each repeat as either transposon, putative microsatellite, LINE, direct repeat, centromeric repeat, etc.

If you have any questions or confirmatory or contradictory evidence concerning the annotation of F9H3, please direct email to Larry Parnell at [parnell@cshl.org](mailto:parnell@cshl.org).

BAC F9H3 was sequenced as part of the Arabidopsis genome sequencing effort of the Cold Spring Harbor Consortium. For additional information, please see <http://www.cshl.org/Arabidopsis>. Fingerprint data indicate F9H3 overlaps with F4C21 toward the telomere and with T5L23 toward the centromere. F9H3 contains marker m1233 at 18.8 cM on the Lister & Dean RI map.

**FEATURES**

source	Location/Qualifiers
	1..95519
	/organism="Arabidopsis thaliana"
	/mol_type="genomic DNA"
	/cultivar="Columbia"
	/db_xref="taxon:3702"
	/chromosome="IV"
	/map="18.8 cM"
	/clone="BAC F9H3"
	/notes="hybridizing YAC is C1C8B1"
misc_feature	1..31783
	/notes="overlap with T5L23, GenBank accession number AC005142, from position 26 to 31808; Region: overlap with T5L23"
gene	851..1719
	/gene="F9H3.20"
	/pseudo
	/evidence=not experimental
repeat_region	complement(981..1120)
	/notes="function=unclassified; similar to F6N23, GenBank accession number AF058919"
repeat_region	rpt_type=dispersed
	3173..3969
	/notes="function=putative transposon; similar to FCA1, GenBank accession number Z97336; moderately similar to reverse transcriptase gene of DM15 non-LTR retrotransposon, GenBank accession number L47193"
repeat_region	rpt_type=dispersed
	5889..6303
	/notes="function=chloroplast_repeat"
	/rpt_type=dispersed

complement(6653..12091)  
 /gene="F9H3.19"  
 /notes="encodes putative calcium channel; genomic copy of EST W43766; genomic copy of EST T43251; EST T43251 apparently encodes intronic sequence; possible alternate exon 1 from position 12772 to 12846 is predicted by a NetPlantGene splice donor but has no EST coverage; gene model last edited on 27 Jan 99"  
 /evidence=not experimental

CDS  
 complement(join(6653..6746,6833..6870,6991..7062,7175..7261,7436..7526,7682..7746,7985..8153,8229..8292,8380..8487,8757..8814,8891..9034,9127..9192,9311..9383,9530..9591,9745..9822,9929..10034,10127..10293,10391..10451,10557..10580,10644..10774,10890..10930,11292..11403,11500..11643,11972..12091))  
 /gene="F9H3.19"  
 /note="functional catalog ID=07.01"  
 /codon\_start=1  
 /evidence=not experimental  
 /product="putative calcium channel"  
 /protein\_id="AAD11598.1"  
 /db\_xref="GI:4206210"  
 /translation="MEDPLIGRSLGGGTDVRRSEATHTGTPFQKAALVDLAEDGIGLPVEILDSSGSEASIIYFTRDLIWSLNFALLFLNFEQPLWCNKPKPCKDRDYVIGELPYLTNAESIYEASRIYFMTSLNLVKVACVILFVDVLFLYLSPLAFDPLPFIAPYVRVIFILSIQRQCADHVMELRDLTLVLSGMLGYLNLALWMLPLFASWIAFMFEDTQGLTVFTSYGATLYQMFTLTSSNPDMVIAIKSSRVSVPFVLVGLVYFVNLILAVDYDFKQALQVSGMDQMKRMLEKAGLIDSDKNGEIDKQCIKLPQNLINRYLPSKKEEFGILFDELDDTRDKINDEFADCCALRFPKEEVLFFHFPQIYHSALQAFVRSFNFGVIAIFILINFIADVVEVTEIDIEESAOQWQVAFVFGWIVLEMAKLIYTYGFENYWRGANRFDLTVTWIVIGETATFITPONTFFSGNGEYIRLLRLMNVORVRAPIATFTILPSLMEYLTGTFVCLCTYSIGVQVFGGLVAGNKLFEELAEADYLLFNFDYNGMVTFLNLVGMNQVMESYKDLTGTMWSTIYVSFYVITILLILNLAVAFLEAFTELDLEEBEKCQCGDSEKRRRSAGSKSRQVRDILLHMLGDELSPKCSSTDT"  
 complement(13149..18831)  
 /gene="F9H3.18"  
 /notes="encodes putative glucan synthase component; gene model last edited on 5 Jan 99"  
 /evidence=not experimental

CDS  
 complement(join(13149..14597,14778..16800,16961..18831))  
 /notes="similar to 1,3-beta glucan synthase; functional catalog ID=01.05.99; functional catalog ID=09.01"  
 /codon\_start=1  
 /evidence=not experimental  
 /product="putative glucan synthase component"  
 /protein\_id="AAD11597.1"  
 /db\_xref="GI:4206209"  
 /translation="MSLRHRTVPQTGRPLAAEAVGIEBEPYNIIPVNNLADHPSLRFEVRAAAALKTGDLRRPVQWRSYDVLDDMLALFTGFKDKNVRNREHVLHLAQAQRLSPPPDNIDSLDSAVRRFRKRLANYSWCYLGKKSNIWISDRNPBSRELVLVGLYLWGEAANRFPPECICYIFHNMSLNKLCLEDDENTGQPYLPSUGENAFLVGVVPIYDTIQAEIDESKNGTVAHCKRNYDDINEYFTWDFCKLWPLDLSGNPFPSKSGVGTGFVETFFYLSRDLWMLALFLQAAIIAWEKPKDTSVTRQWNAKADGVORLLVFLTWGSMRLLQALDAASOYPLVSRTRKRRHPRMLMKVIAAAMVIAFTVLTNIWKQKRODWSNAATTKIYQYLAAGVAFLEILALALPIWMRNPLEETNKKFFALTWNFQGSFVGRGLRGLVDNIKYSTFWIATKFTFSFPLQVPMIKPSKLLNWKDVIDYEMHQFQDSNRFSVALLPLVLIYDQIDQIWAIVISSIVGAVGLFDHLGEIRDMQRLRFQFASAIQFNLMPEQLNARGFKGDKGHRKLRVYGFGRFPFKLESQVANKFALINWEIIILAFREEDIVSREVELLELPKNSMDVTVIRWPCFLLCNELLALSOARELIDAPDKLWHKICKNEYRCACVAAVDSIKHLLSIIKVDTEHSIIITVFQIINOSIQSEQFTKTRVDLLPKIYETLOKLVLVNDDETDGRVNVNLSQSLYEIATQFFIEKTKTEQSLNEGLTRDPASKLLFQNALRPLDASNEDPYQVRRLHTILTRDSMHSVPVNLREARRIAFFSNLFMMNMPHAPQVEKQMSAVLTFRYSEEVYVSKQLRNETEDSTLYLQTIYADEWKFMRKREGIKTGKIDPFWTKLRDLWASVYGOTLARTVRGMYYRALLMLAFALDSAGMKTREGAQELGVSQVNLQCELGQSDGFVSENDSSLSRASSSVYTLKGYEYGTALMKFTVYVACQIYGSQAKKEPOAEILYLKQNEALRIAYDVEPAGGETDYVSLVKYDHOLEKVEIPLVKLPQPVKLGCKPENQNHAMIFTRGDVAQTTIDMQDSYFEALKMRNLKQNYNHYHGRKPTILGVREHIFTGSSVSLAWFMSAQETSFTLQGRVLANPLKMRNHYHDPVDFRWFLSRGGISKASRVINISDIFAGFNCTLRGNVTHHEYIQVGRKDVGLNQISNFEAKVASGNGEQVLSRDVYRLGHRLDFFRMLSPFTYTTVGFNTMMVILTVYAFWGRVYLALSGV

gene

CDS

gene

CDS

EKSLADSTDTNNAALGVILNQOFLIQLGLFTALPMIVENSLBEGFLALLWFIWMQIQ LSAFTFSMGTRAHFTGRITLHGGAKYRATGRGFVVEHKGFENTRLRLVARSHFVKAI EGLNLIIVYASHSPIAKSLIYIAMTITSMFLVSIWMAFPFVFNBPGLMKTVYDFE DFNMTWYQCRISTKSEQWKEWMEODHLRNTGKAGLFVBIILVLRFFQYQYIVY QLKIANGSTSLFVLPFSWYIYFAIFLVIOVARDKYSAKAHIRVLRVLFQYLLIVAI LVTVALLERTHSRFDIFTSLLAIFPTGWLILLIAOTQRKWLKNYTFIWNVUSVARM YDLFCILLMVPVAFLSWMPGQSMQTRILFNEAFSRGLRIMQIVTGKSKGDV" complement (20861. .22302) /genes="F9H3.17" /notes="encodes hypothetical protein; gene model last edited on 5 Jan 99" /evidence=not experimental complement (join(20861. .21094,21937. .22039,22145. .22302)) /genes="F9H3.17"									
gene									
CDS									
Query Match	59.4%	Score	571.6;	DB	8;	Length	95519;		
Best Local Similarity	92.2%	Pred. No.	1.7e-79;						
Matches	636;	Conservative	0;	Mismatches	49;	Indels	5;	Gaps	3;
QY	1	TGTGTTTCTCAGAAATAGACGCAAAATATTATATAAAGCATGCAATTCCTTATAGATCG	60						
Db	23202	TATGTTTCTCTGAAATAACACGATATATTATTATAAAGCATGCAATTCCTTATAGATCG	23261						
QY	61	CGAAG-TTTAAAAAACAATAGAAATTTGTACAAATATTACATCGGTTTTTATTGGATAAC	119						
Db	23262	CGAAGTTTTTAAAAACAATAGAAATTTGTACAAATATTACATGAGTTTTTATTGGATAAC	23321						
QY	120	ATGACAAATATTATTATTCTCATGAGTTTTTATTGGATAGCATGACAAATATTAATATA	179						
Db	23322	ATGACAAATATTATTATTATATACATGAATTTTCTGTGATAGCATGACAAATATTAATATA	23381						
QY	180	TCAGTGTAAATACATGTTTGTCTTAAATACATGCATTTTAAATCAGACATTTGTT	239						
Db	23382	TCAGTGTAAATACATGTTTGTCTTAAATACATGCATTTTAAATCAGACATTTGTT	23441						
QY	240	TAAAAATCAATCTAATCTTATATACAAACACATTTGACGAAAAATTCAGTGAATAAG	299						
Db	23442	TAAAAATCAATATATATCTTATATACAAACACATTTGATGAAGATTCAGGCAATAAG	23501						
QY	300	AGAAAAATAAGAAATAGAGATAGAGATTTCTATGAAAAAAGAGAGACATGTAG	359						
Db	23502	AGAAAAATAAGAAATAGAGATAGAGATTTCTATGAAAAAAGAGAGACATGTGG	23561						
QY	360	GTGACAAATAAAGAGATATGATATATTTTATGAGGGTGGTGAAGATTAATTTAG	419						
Db	23562	GTGACAAATAAAGAGATATGATATATTTTATGAGGGTGGTGAAGA-TATTCTAG	23620						
QY	420	GAGAGGGAGAGAAATAGAAAAAGAAATGACATGCTGAATCTGAAGAAGATGAATTTG	479						
Db	23621	GAGAGGGAGAGAAATAGAAAAAGAAATGACATGCTGAATCTGAAGAAGATGAATTTG	23680						
QY	480	GTTTAAAGATGAAGAGAGAAAGAGAAAGCGCTAAATAGACTAATTCGCCAAATTT	539						
Db	23681	GTTTAAAGATGAAGAGAGAAAGAGAACTTCATGCTTAAAGTTTCGTAAGAAACATGAAA	23740						
QY	540	GAACAAAGAGAGAGAGAGAAAGAGAAAGCGCTAAATAGACTAATTCGCCAAATTT	599						
Db	23741	AAAAGAA---AAAGAGAGAAAAAGAGAAAAATATAATAGACTAAATATTGCCAAATTT	23797						
QY	600	CTGTAGCGCAAAATACTATTTCGTCCAGGTTATTTTGTGTAATCTTTTGAAGTCAAAA	659						
Db	23798	CTGTAGCAGCAAGTACTATTTCGTCCAAAGTTATTTTGTGTTTATTTGTTTCAAGTCAAAG	23857						
QY	660	GTTTATTCTTTACATATACTCTAAAAATATA	689						
Db	23858	GTTTATTCTTTACATATACTCTATATATA	23887						
RESULT 10									
AC005142/c									
LOCUS									
DEFINITION									

ACCESSION	AC005142								
VERSION	AC005142.2	GI:4263038							
KEYWORDS	HTG.								
SOURCE	Arabidopsis thaliana (thale cress)								
ORGANISM	Arabidopsis thaliana								
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsiis.								
AUTHORS	Zhong,J., Ma,P., Parnell,L.D., Chen,C.-N. and Chen,E.Y.								
TITLE	BAC T5L23 from chromosome IV, position 19 cm								
JOURNAL	Unpublished								
REFERENCE	2 (bases 1 to 116448)								
AUTHORS	Zhong,J., Ma,P., Parnell,L.D., Chen,C.-N. and Chen,E.Y.								
TITLE	Direct Submission								
JOURNAL	Submitted (19-JUN-1998) Applied Biosystems Division of Perkin-Elmer, 850 Lincoln Centre Dr., Foster City, CA 94404, USA								
REFERENCE	3 (bases 1 to 116448)								
AUTHORS	Parnell,L.D. and Chen,E.Y.								
TITLE	Direct Submission								
JOURNAL	Submitted (17-FEB-1999) Applied Biosystems Division, Perkin Elmer Corporation, 850 Lincoln Centre Dr., Foster City, CA 94404								
REMARK	Arabidopsis thaliana BAC T5L23 from chromosome IV near 19 cm								
COMMENT	On Feb 22, 1999 this sequence version replaced gi:3241935 BAC T5L23 maps to near 19 cm on the Lister & Dean RI map and is assigned to YAC C1C8B1. Position 1 of T5L23 is oriented toward the telomere and position 116473 is oriented toward the centromere. For more information on the mapping, sequencing and annotation of T5L23, please see <a href="http://www.cshl.org/Arabidopsis/T5L23-titlepage.html">http://www.cshl.org/Arabidopsis/T5L23-titlepage.html</a> . A graphic view of our annotation is also available at this url. Gene models are built with exons predicted by GenScan ( <a href="http://CCR-081.mit.edu/GENSCAN.html">http://CCR-081.mit.edu/GENSCAN.html</a> ), MZEF ( <a href="http://www.cshl.org/genefinder">http://www.cshl.org/genefinder</a> ) and GRAIL ( <a href="http://compbio.ornl.gov/tools/index.shtml">http://compbio.ornl.gov/tools/index.shtml</a> ) and with splice sites predicted by NetPlantGene ( <a href="http://www.cbs.dtu.dk/netgene/cbsnetpgene.html">http://www.cbs.dtu.dk/netgene/cbsnetpgene.html</a> ). Genes are numbered according to the scheme BAC.gene_number. Typically, these numbers progress from 1 upwards as one moves from position 1 of the BAC. Protein sequences encoded by the genes are assigned to a functional category with the aid of similarity searches and comparison to the Prosite ( <a href="http://expasy.hcuge.ch/sprot/prosite.html">http://expasy.hcuge.ch/sprot/prosite.html</a> ) and Pfam ( <a href="http://pfam.wustl.edu/">http://pfam.wustl.edu/</a> ) libraries. A description of these categories can be found at <a href="http://muntjac.mips.biochem.mpg.de/arabi/">http://muntjac.mips.biochem.mpg.de/arabi/</a> . Genomic repeats are typically located by TBLASTX analysis and an attempt is made to classify the function of each repeat as either transposon, putative microsatellite, LINE, direct repeat, centromeric repeat, etc.								
FEATURES									
source	1. .116448 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /cultivar="Columbia" /db_xref="taxon:3702" /chromosome="IV" /map="19 cm" /note="hybridizing YAC is C1C8B1"								
misc_feature	1. .31783 /note="overlap with F9H3, GenBank accession number AF071527, from position 1 to 31783; Region: overlap with F9H3" complement (2025. .3495) /gene="T5L23.1"								
gene	/note="encodes putative M-type thioredoxin; identical to F9H3.15, GenBank accession number AF071527; genomic copy of EST AA042639; genomic copy of EST AA650755; gene model last edited on 5 Jan 99" /evidence=not_experimental								

CDS	complement(join(2025..2341,3252..3495)) /gene="T5L23.1" /note="similar to A. thaliana protein F21B7.7, GenBank accession number AC002560; M-type thioredoxin is also present in cyanobacteria and chloroplasts; functional catalog ID=02" /codon_start=1 /evidence=not_experimental /product="putative M-type thioredoxin" /protein_id="AAD15308.1" /db_xref="GI:4263039" /translation="MAAFTCTSRPPISLRSETRIYSSSPASSLSRRMPAVLPESGG LRLRLSPASLTSIHQPRVSRURRAVCEAQETTDIQQVNDSTMDSLVLKATGPVV VDWAPWCGPKMIDPLVNDLAQHTGKIRFYKLNLTDESPNTFQGVRSIPTIMIFV GSKKDTIIGAVPKTTLTSSLDKFLP" complement(4139..4163) /note="function=putative_microsatellite" /rpt_type=tandem /rpt_unit="a" /rpt_start=1 /rpt_end=5120 /note="function=LTR_transposon" /rpt_type=dispersed 5120..5124 /note="function-terminal_flanking_direct_repeat; cognate A of flanking direct repeat of LTR transposon from position 5145 to 6873; cognate B residues from position 6869 to 6873"
repeat_region	
repeat_region	
repeat_region	
LTR	5125..5128 /note="cognate A of LTR flanking transposon from position 5145 to 6873; cognate B residues from 6665 to 6868" 5508..6632 /gene="T5L23.2" /note="encodes putative reverse transcriptase; likely functions as an element of LTR transposon from position 5145 to 6873; gene model last edited on 5 Jan 99" /evidence=not_experimental 5508..6632 /gene="T5L23.2" /note="likely element of LTR transposon from position 5145 to 6873"
gene	
CDS	/codon_start=1 /evidence=not_experimental /product="putative reverse transcriptase" /protein_id="AAD15309.1" /db_xref="GI:4263040" /translation="MGRNKGKVRSSPDLIVIASKIHQSGSSARAPGSHHAA IHQPAQYSSRRHFEVSDPNSSPHLVSSDHPGLVLAPELDDGNSYGTWIIAMTTS IEAKNLGVDGSIKPKDDDDPYCKIWRRCNSMVKSLNSVKEIYTSILYPTAAA IMKDLVTRFKSLPLRYKLQQTIIHSLRQGNLDSLSSYHTRKQLWBELTSLQAPRTV EDLIERETNRVIDFLMGLNDCYDVRSQILMKKTLPSLSEVFNMIDQDEIORSARIS TTPQMTSSVFAVSNQSSQVLNGDTYQKRPVCTYCSRPGHVEDTCYKKGYPFSK SKQIANTELIVDPFTKLLYPSHFQRLSKMGLLNIFVPS" 6640..6843 /note="cognate B of LTR flanking transposon from position 5145 to 6873; cognate A residues from 5150 to 5353" 6844..6848 /note="function-terminal_flanking_direct_repeat; cognate B of flanking direct repeat of LTR transposon from position 5145 to 6873; cognate A residues from position 5145 to 5149" /rpt_type=flanking complement(7885..7904) /note="function=putative_microsatellite" /rpt_type=tandem /rpt_unit="at" complement(7905..8818) /note="function=unclassified; similar to MXI10, GenBank accession number AB005248" /rpt_type=dispersed 9482..10923 /gene="T5L23.3" /note="encodes hypothetical protein; identical to F9H3.17,
LTR	
repeat_region	
repeat_region	
repeat_region	
gene	

CDS	GenBank accession number AF071527; gene model last edited on 5 Jan 99" /evidence=not_experimental join(9482..9639,9745..9847,10690..10923) /gene="T5L23.3" /note="identical to F9H3.17, GenBank accession number AF071527; similar to F21B7.5, GenBank accession number AC002560; functional catalog ID=99" /codon_start=1 /evidence=not_experimental /product="hypothetical protein" /protein_id="AAD15310.1" /db_xref="GI:4263041" /translation="MWKLTKRIGGLVLRLAFAAGLAALIVMITSRRASFLAISLEA KYTDMAFKYFVIANAVVSVFLVFLPKESLWKFVVVLDLVMTLTSSLSAALA VAQVGKGNANAGWLPICGQVPKFCQITGALITAGFVALVLYLLVLSLHAVVDPFL LOKS" 12953..18635 /gene="T5L23.4" /note="encodes putative glucan synthase component; identical to F9H3.18, GenBank accession number AF071527; gene model last edited on 5 Jan 99" /evidence=not_experimental join(12953..14823,14984..17006,17187..18635) /gene="T5L23.4" /note="similar to 1.3-beta glucan synthase; identical to F9H3.18, GenBank accession number AF071527; functional catalog ID=01.05.99; functional catalog ID=09.01" /codon_start=1 /evidence=not_experimental /product="putative glucan synthase component" /protein_id="AAD15311.1" /db_xref="GI:4263042" /translation="MSLRRTVPPTGRPLAAEAVGIEEPYNIIPVNNLADHPSLR FPEVRAAAALKTVGDLRPPYQVQWNSHYDLDLWALFFGQKDNVRNQEHVWLHLA NAQRUSPPDNDISLSDSAVFRFRKLLANYSWCYSYLGKSNITSDRNPSSRREL LVYGLYLLWGEAANLRFMPECIYIFHNNASELNICKDCLDNTGQPYLPSSLGEN AFLTGVPKPIYDTIQAEIDESKNGTVAHCKRWNYDDINFEYTRDRCFQSKLWPLDLGS NPFKSGKSGVTGTFEERTFFYLRSFDRLLWMLDALFQAAIIVAEKPKDTSVTR QLWNLAKARDVOVRLLTVELTWSCMBLLQALDAAQYPLVSRBTKHFRFRLMKVIA MNVIAVFTLYINIKQKQKQRONNAATKIYQFLYAVGFLVPELALALFIIPW ARNFBLEETNWKIYFFALTWPGQKSFVGRLEGLVDNIKYSTFIWFLATKFTFSYPL IGVKPMIKPSKLLNKLNDVDYEHQFQSDNSRFSVALLMPEVLIIYMDIOWIYIYSS IVQAVVGLFDHLGEIRDMLQRFQFASAIQFNLMPERQLLNARGFGNKDGIIHR LKLYGFGRRPFKKLESNOVEANKFALIWNEIILAFREEDIVSDREVELLELPKNSMDV TWIRWFCFLCNELLALSOARELIDAPKMLWHKICKNEYRCANVEADYSIKHLLD SIKIDTEHSIITVFQIINISQSEQFTKTFRVDLLPKIYETLQKLVGLVNDDEETD SGRVNVLSLYEIAIATROFFIEKKTEQSNEGLPRDPASKLLFQNAIRLPDASNE PYQVRRRLHTILTSDSMHSVPVNLERRIARFNSLPMNMPHAPQVKKMAFVSILT PYVSESVVYSKEQLRNETEDGISTLYLQTIYADEWKNFKERMHREGIKTDSLELWTK LRDRLWASVYRGQTLARTVYRGMVYVYRALKMLAFLDSASEMDIREGAQLGSLVRNLOG
gene	
CDS	
Query Match	59.4%; Score 571.6; DB 8; Length 116448;
Best Local Similarity	92.2%; Pred. No. 1.6e-79;
Matches	636; Conservative 0; Mismatches 49; Indels 5; Gaps 3;
Qy	1 TGTGTTTCTCAGAAATAGCACGAAATATTTATATAAAGCATGCAATTCCTCTATAGATCG 60
Db	8582 TAATGTTTCTGAAATAACACGATATATTTATTTAAAGCATGCAATTCCTGTTATACATCG 8523
Qy	61 CGAAG-TTTAAAAAACATAGAAATTTGTTTACAAATATTACATGGGTTTTTATTGGATAAC 119
Db	8522 CGAAGTTTTTAAAAACATAGAAATTTGTTTACAAATATTACATGAGTTTTTATTGGATAAC 8463
Qy	120 ATGACAAATATTATTATTTTTCATGAGTTTTTATTTGGATAGCATGACAAATATTATATATA 179
Db	8462 ATGACAAATATTATTATATTACATGAATTTTCTTGGATAGCATGACAAATATTATATATA 8403
Qy	180 TCAGTGTTAATAACATGTTTTTGTCTTAAATACATGCATTTTAAATACAGACATTTGTT 239
Db	8402 TCAGTGTTAATAACATGTTTTTGTCTTAAATACATGCATTTTAAATACAGACATTTGTT 8343
Qy	240 TTTAAATCAAACTAACTCTTATATACAAACGACATTTGACGGAATTCAGGTAAAAAG 299



```

CDS
complement(join(10335..10525,10983..11078,11590..12133,
12345..12500,12627..12772,12871..13501))
/gene="AT4g03490"
/notes="contains similarity to mammalian ankyrins
similarity to: similar to A. thaliana hypothetical protein
F21B7.8, GenBank accession number AC002560"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB77834.1"
/db_xref="GI:7270672"
/db_xref="UniProt/TREMBL:O9ZT77"
/translation="MGKQEFKMEENEPVLDQVTFQGTWTLHAAIYCHDHLVRR
ILAYELNLRKRGKLCNFPVFSFYOTLLVRNYGDLALHVAAGHKLVLGLLI
DCLRQPDITWIGSQWIGNIFRVSNNDGNTALHSLKGNHVSLSQLVREDRST
CFLLKEDVPLMAAEGYVSLVHMLRGDASFVGSKVLCVCAVKSQNDIITAVLE
SDSLVESRDEGRTPUATASGIDYGVQHLTRFASSTQGNQVHLVAASKGNARAV
GYLLKSDVKRLNEODIEGNTPLHLASSNHKPVWLIWMALVAAGCTTRPVHLRAD
IPGLTDEDLILKHDRVNTLLVATLVATMAFAGLSVPLGYNSTFEKSNVKSVE
ESAPFAPVINCSTAVYTAIVSTVALIGTQADLKCMUTTFKFIPLPLGFSIIAMSLAF
VAGLYLVGHYHMLAIFVLASGGFYLMAILLIIPYASPTFTLSRSLNSLVQMSKE
DVDSVNLVPAPTEELAIKAIQVNRICIQGNIGVTSLSYLLKHVAPINVLLHLKLN
HYSLVFSGLYDGPKTYSQL"
complement(10335..10525)
/gene="AT4g03490"
/number=1
complement(10526..10982)
/gene="AT4g03490"
/number=1
complement(10983..11078)
/gene="AT4g03490"
/number=2
complement(11079..11589)
/gene="AT4g03490"
/number=2
complement(11590..12133)
/gene="AT4g03490"
/number=3
complement(12134..12344)
/gene="AT4g03490"
/number=3
complement(12345..12500)
/gene="AT4g03490"
/number=4
complement(12501..12626)
/gene="AT4g03490"
/number=4
complement(12627..12772)
/gene="AT4g03490"
/number=5
complement(12773..12870)
/gene="AT4g03490"
/number=5
complement(12871..13501)
/gene="AT4g03490"
/number=6
13514..15185
/notes="marker mi233"
14170..17288
/gene="AT4g03500"
join(14170..14204,15103..15848,15939..16477,16650..17288)
/gene="AT4g03500"
/notes="hypothetical protein with some similarity to
ankyrin
similarity to"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB77835.1"
/db_xref="GI:7270673"
/db_xref="UniProt/TREMBL:O9ZT78"
/translation="MKNESRSTSEISKLGIKIFSGVGSFMSGRPDVHLDRVAENHQAIPT
NNVRSPLSLDLSFDTSKPMPTMAAVRAGKENVLRNSNYSISVAPTLVNORG
NTLHLAASGHVSLVRYIIQKCPGLLLKMMGEVALHAAAGHLDVWNLDFTN
DISCTNLPVAKRIYFAKNKNQDILHVALKGHEWASVLSVSAKSLSFVANRDGFSF

```

```

LYLAIBAGHTSLVTTMCHGTNELSSKVGGRSIVHAAKANKRKDILDALLSKDASLNL
KDEGRSLSFGASIGYQGSYLFKGNRKVYVDDDDGLFPPTHMAKYGHVQLEEL
KHCEAIELDRDQONILHAAKYKLVKIFLSCCKNKKKLNEODVNGTPLH
LATINHPKVSMFTDHRVLDKKNYIGFTALDVAEENIDSSYIVHORI.TTMALINA
GAPKSTPITENLRSPKKGKORVNTLMVLATLVATMTTGTATLFGGYNDSFP
HIGMAVAKRTAFQVFLVCDTLAMYSIIITIVAIWAQGLDLSIILKAFNIALPFLGL
ALTSMSIAFMAGTYVAVSHPLGLGYFVLGIIIFLLVLLLLLPVYSPYAHAPLLRH
IFYYPYFLKLLAAAGDNKINVDVYTASDE"
14170..14204
/gene="AT4g03500"
/number=1
14205..15102
/gene="AT4g03500"
/number=1
15103..15848
Query Match 59.4%; Score 571.6; DB 8; Length 159629;
Best Local Similarity 92.2%; Pred. No. 1.5e-79;
Matches 636; Conservative 0; Mismatches 49; Indels 5; Gaps 3;
QY 1. TGTGTTTCTCAGAAAATAGCACGAAATATTTATAAAAAGCATGCAATTTCTTATAGATCG 60
DB 29859 TATGTTTCTCTGAATAACACGATATATTTATTAAGACATGCAATTTCTGTTATACATCG 29880
QY 61 CGAAG-TTTAAAAAACATATAGAAATTTGTACAAATTTACATGGGTTTTTATTGGATAAC 119
DB 29799 CGAAGTTTTTAAAAACATATAGAAATTTGTACAAATTTACATGAGTTTTTATTGGATAAC 29740
QY 120 ATGACAAATATTTATTTATTTTCATGAGTTTTTATTTGGATAGCATGACAAATATTAATATA 179
DB 29739 ATGACAAATATTTATATATTTTACATGAATTTTGTGGATAGCATGACAAATATTAATATA 29680
QY 180 TCAAGTGTAAATACATGTTTTTGTCTTAAATACATCATTTTAAAAATCAGACATTTGTT 239
DB 29679 TCAAGTGTAAATACATGTTTTTGTCTTAAATACATCATTTTAAAAATCAGACATTTGTT 29620
QY 240 TTAATAATCAAACTCTAATCTCTTATATCAACACGACATTTGACGAAAAATTCAGGTAAGAAAG 299
DB 29619 TTAATAATCAAACTCTAATCTCTTATATCAACACGACATTTGATGAAGAAATTCAGGCAAAAG 29560
QY 300 AGAAAATAAGAAATGAGAGATAGAGATTTCTATGAAAAAGAAAGAGAGAGACATGTAG 359
DB 29559 AGAAAATAAGAAATGAGAGATAGAGATTTCTATGAAAAAGAAAGAGAGAGACATGTGG 29500
QY 360 GTCAACAAAAATAAGAGATATGATATATTTTATGAGAGGTGGTCAAGATTTATTTTAG 419
DB 29499 GTCAACAAAAATAAGAGATATGATATATTTTATGAGAGGTGGTCAAGATTTATTTCTAG 29441
QY 420 GAGAGGGAGAGAGAAATAGAAAAAGAAATGACATGGTGAATCTGAAGAGAGATCAATTTGT 479
DB 29440 GAGAGGGAGAGAGAAATGAGAAAAAGAAATGACATGGTGAATTTGAAGAGAGATCAATTTGT 29381
QY 480 GTTAAAGATCAAGAGAGAGAAAGAGAACTCCATGGCTAAAGTCTCGTAAAGAGATGAAAAA 539
DB 29380 GTTAAAGATCAAGAGAGAGAAAGAGAACTTCATGGCTAAAGTTCGTAAAGAGACATGAAAAA 29321
QY 540 GAAACAAAAAGAGAGAGAGAAAGAGAAAGCGCTAAATATAGACTAACTATTATGCCAAAAATTT 599
DB 29320 AAAAGAA---AAAGAGAAAAAGAGAAAGAAATAGACTAAATATTATGCCAAAAATTT 29264
QY 600 CTGTAGCCGACAAATACTATTTGGTCCAAAGTTATTTTGTGTATTTCTTTTGAAGTCAAAA 659
DB 29263 CTGTAGCAGACAAAGTACTATTGTGTCANAGTTATTTTGTGTTATTTGTTTCAAGTCAAG 29204
QY 660 GTTATTTCTTACATATATCTCTAAAAATATA 689
DB 29203 GTTATTTCTTATATATCTCTATATATA 29174

```

```

RESULT 12
AC007203/c
LOCUS
DEFINITION Arabidopsis thaliana chromosome 95190 bp DNA linear PLN 30-OCT-2002

```



```

complete sequence.
AC007203 AC007203.3 GI:4996903
VERSION HTG.
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;
Rosids; eutrosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 95190)
Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altati,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
Unpublished
2 (bases 1 to 95190)
Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altati,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
Direct Submission
Submitted (03-APR-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
3 (bases 1 to 95190)
Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altati,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
Direct Submission
Submitted (04-JUN-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
4 (bases 1 to 95190)
Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altati,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
Direct Submission
Submitted (16-JUN-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
On Jun 4, 1999 this sequence version replaced gi:4678187.
Bases 1-6718 of clone T10P12 overlap with bases 122,554-129,271 of
'IGF' BAC clone FI121. gb|AC005687
e-mail for correspondence: arabes@sequence.stanford.edu
Genes with similarity to proteins in the databases are described as
'putative', '-like' or 'similar to'. Genes that have EST
similarity but no significant protein similarity are described as
'unknown proteins'. Genes that are annotated based only on gene
prediction software are described as 'hypothetical proteins'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/section/index.html), GENSCAN (Chris Burge,
http://gnomic.stanford.edu/~chris/GENSCANW.html), Fexa (V.Solovyev
& A.Salamov, Sanger Centre, http://genomic.sanger.ac.uk/),
NetPlantGene (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark,http://www.cbs.dtu.dk/NetPlantGene.html) and
eMotif(Nevill-Wanning, C.G. Wu, T.D. & Brutlag, D.L.,
http://motif.stanford.edu/projects.html).
FEATURES
Location/Qualifiers
1. .95190
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="I"
/clone="T10P12"
1. .6718
misc_feature

```

```

/note="overlap with bases 122,554-129,271 of 'IGF' BAC
clone FI121. gb|AC005687"
repeat_region
9474. .10628
/rpt_family="LTR"
gene
complement(19003. .23764)
/gene="T10P12.1"
CDS
complement(join(19003. .19302,20071. .20661,20728. .20928,
21074. .21678,21933. .22099,22247. .22640,22931. .23131,
23362. .23764))
/gene="T10P12.1"
/note="Similar to reverse transcriptase; Similar to reverse
transcriptase, part of putative retrotransposon"
/codon_start=1
/db_xref="GI:5080760"
/db_xref="AA039270.1"
/translat="MDMIFNVGDKVRRKIENSKTAAEAWGTDILYLKSLPNRVYL
OLKVTNYRMNSNTLEENVDPEKMSIDLNLIQIOVPDEVAQVAILIUSALPDSODM
LKETLYKGREGIKLDDVISAISKELELSDSTSEVNLGSGKDHGEWILIDTECSFHM
TRKXYHIDFEAKSGKVRPNNSPLEVGIGIKGRTFNQDCTSCFNDNRISKIKFCENF
VNGTKRTYSFGTAQVHTKELDYVHSDUMGSSNPVPSLSKQCFISFTDWDGRKXWIN
FNTKDEAFTFTKWKVETOSERLKKLRTDNGLESCNQKDFGFKKEEIVRHRTS
PSSAENKIPLELWTSVIPSLGLELRFVGVHSGEKLDPKAKKGVFVPGSPED
TSSEENKTRQIARDRPRROVLPSRLKDYEMDEVLDEIAGVAYLITEDGGPEPAS
FOEALQSDSGKWLAEDEIOSLIKNTWLVARNKSKQPIKCKWFKRAGIAGVE
KRFKARLVTKGYSQKEGIYFQEIFSPVVKYSIRLLSIVFHLNMQMFKTAFL
KRYDEMIFMDQEGYVHEKYPRIYLLYVDDILIASKKKKHCELXKLVSEFEM
KOLGDAKILGMEIIRKDEGGTLISQEGYLLKLRPATVEYKSGKNVTRPVRSAY
GSLMYSMITGRPOLSAHSLGLCMFMRKPLKDHQAVKWLIRLCIVTMDRKLVCYKGG
LALGVCDSYAAADKEKRSTGVTFTGGNTISWRSSLOKVALSSTKAEYMALTDA
TNETWMLKGLVNELOFTORSVNIHCDOSVIALKANVYHEHTKHIDVKYHFTRELAK
ANNEKITLKGGSLEEEKQVLANKKEYDDAEATYNNKLVLKELKIPPPVFNLIIP
NSLGNRSAGLEAGVGVTDTHASSNLREDEMCPSEBEPAY"
29564. .30696
/rpt_family="LTR"
/gene="T10P12.2"
complement(join(31637. .32634,33345. .33502,33679. .33916,
34104. .34307,34383. .34539,34828. .35319))
/gene="T10P12.2"
/note="Similar to reverse-transcriptases; protein is
flanked by LTRs, potential retrotransposon"
/codon_start=1
/protein_id="AAD39271.1"
/db_xref="GI:5080761"
/translat="WMEQLRIGTCYFGSGTSPEADSWRSRVERNFSGSKCSVEYLV
DLAVHFLGEGDAHLWRSVTAKRQEDMAWDFVAFBNAKYCQEAALDRMEARFLELTQ
GSRVREYDREPNRLLYAGQMEDDQAQMRFLGLRDPWRCVRCSOYATKAALVE
TAAELMVAVVQPAVQOQGAOVQOQNVQQLAHGAA PQGYTTREIGSTSKRAITGL
LAHEVCVGTLLVDGVEAHVLFDSGASHCEITPKSASRGNIRGDPGEQSGSVKAGQGF
LAEVGRARVKEAKMEKGEAYLMTISMPESQGVAVSNIRVYQDEFDFPSQLGLP
PWSMDFPTIVLPEPTALLSKHPTGMLQQRWIDELDLQURGATCFSKIDLTSGVHPII
ASADIRKTAFTRYGFHFVTFPFAELDAIWIIVDRLTAKSAHLAIRKTDGAALIT
KYYVSEIVKMGVPSVIVSDRSKFTSAFRAQEMTKVQMSYAHVHPQTQGSERT
IOTLDEMLRMVLDMGHMADOLSLIEFAYNNYQASIGMTFVILVGPCHTOLCWT
QVGRSIVGANYVQETTERIGVILKLMKEVQDRHSYADKRRRELEFEFGERVYLKMA
KURGNRSISETKLSFRYMGPPRVVFAIRLELTDVMAFRAHFVHVSIMRKLCH
MDDEVLAKIPLEDLQNMTEARPGRVLRRIKELRRKKIPIKVLWDCDGVTEETWEP
EYRMKARF"
35970. .37136
/rpt_family="LTR"
/gene="T10P12.3"
37245. .42330
CDS
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(37249. .37347,37714. .38098,38266. .38532,38897. .39097,
39215. .39508,39603. .40889,41069. .42330)
/gene="T10P12.3"
/note="Similar to retrotransposon polyprotein"
/codon_start=1
/protein_id="AAD39272.1"
/db_xref="GI:5080762"
37245. .42330
/gene="T10P12.3"
join(3724
```



OMVSRGKNNVAAATYNTKSYTKPFLITSGAPVAHRSKLLKCADNNLVLMLGNSVSH  
CSGVCKSLVFLWLGVEFGNFMSELGVVVDVVLGVEMLETLAVPKLVEIHESDLTEGS  
LSLSSGIALMEMPLVVEVLESSEPTIFDIHLELPPFLEHSHLSLKPGVSSILVRSVRY  
PHATNVVLEKMEVEMLVGTGIIIRPTIIPKFPPIVIELLDLHGMIFSKILDCSGYH  
QRMKPEDLEKTAFTLEHGYBFLVMSFGLTNAPATPFQSSMNRISKKLLRKFVLFVFD  
HLIIVSPSESHLREVLQDLNKLKYNMKCSFGVQVVEYLGHIIISABGVVDS  
TKTEAHNNQVFNLSKQLRGLTGYIKFVRGYSIAKLTLLKRDKFLWLTWQ  
DAFEVQKAMILAPVFPDFKVFLETDVFGTGVGAVLMOEKRIITVPSHDLTPRE  
OLKSAVERELMAVMDLVKHYLIGRKFVHTDORSLSKFLLEKEVNEYORMLTRL  
LQCDMDIVVYKPGVANKAAGLSRIQHPVYALLALTVPVYIQLDLYLDEDEHIQS  
VIKQHDNLQSSHYQLNGLRWYKQLLVPFKHSSFIPLILQETVWEDISVDPIE  
GUPTSGGVNVLIVVDRLSKAVYFLGKHLFKAIDVANKPSEVVKLYGPPKSI VSDR  
DRIFGAVKDMFKLSQTLQHSYSYRPTQDQTEVNLNRCLETYLRCTSSHPRAWHK  
YLAWEFSNTAYHSALKTSPKVVYVGRPPSLVRPELGSTONWDLVQLKERDLME  
HLRANLQAEITMKRNMADKRDVFAVDGWHLLKLOPYRQRTVVRPCVKLAAPFG  
PREI VERIGNAVRINLPASSKTHNVNSQLKVFGEQVQVQATNPPLKVEDEFLP  
EKVIDISFTDKNGKFLVQDRGVSNSWMPYKVFQVFPQFKLEDKLI FLARSIDM  
IHEAYVRQTRGPHURVEEGEDVAVTPPSLTIRNSAKDPGFQVLUGQFSSGHLPTRC  
TA"

gene complement (47812..49281)  
/genes="T10P12.4"  
CDS complement (join(47812..47877,48046..48223,48314..48405,  
48708..48828,49010..49281))  
/genes="T10P12.4"

/note="Hypothetical protein"  
/codon\_start=1  
/protein\_id="AAD39273.1"  
/db\_xref="GI:5080763"  
/translation="MAISLATAYISPCPTPESSNSASPSRTLSVLRPLSQIRRRFSVQ  
SPSSSTRFAPLTVSSFTNYLFRFDAIDFETKFNVLVLTSSSFVEANRLADGLAN  
YAFSPPLHUESPDVSSILLADVAGVRAAKQTFNSFDLQNSDKPVLVDVFA  
TWGCPQLMVPILNEVSETLKDIIVAVKIDTEKYPSLANKYQIEALPTILFKDGLW  
DRFEGALPANOLVERTENSLOVKQ"  
complement (50976..52602)  
/genes="T10P12.5"  
CDS complement (join(50976..51522,51912..52404,52575..52602))  
/genes="T10P12.5"  
/note="Hypothetical protein"  
/codon\_start=1  
/protein\_id="AAD39277.1"  
/db\_xref="GI:5080767"

/translation="MCLAQMTTISQPLVRYLGLPLMTKAMTADHYLPLIEKIRKISS  
WTRFLSYCGRQLIKSLVMSITNFWSSAFPLPGNCKEIERLSAFIWSGPDLLKTHN  
AKIANGKCLPCEGLGLRPLKEINTVCGLLIWRLLASOTSLMGOWQVOTYLIRNN  
FWAIKASSYQGSWMCMVPOATPKFAPIWLGWHNRLSTGDRMKNGOADSCTVFCOD  
PLETRDLHFCHYANQIWEIIAKGFMGVQYTSNWDQLASLAGTSLEPLFCFPSCML  
PRPPTQRESEITVGMKIQBSTNWRKPKQYETESPFSAITREKNMRTSSVFGS  
PLEANLSLTENFLFLKCCCT"  
complement (60126..62065)  
/genes="T10P12.6"  
CDS complement (join(60126..60521,60607..60800,60875..61272,  
61788..62065))

Query Match 42.3%; Score 406.8; DB 8; Length 95190;  
Best Local Similarity 79.2%; Pred. No. 4.7e-54;  
Matches 588; Conservative 0; Mismatches 82; Indels 72; Gaps 6;

QY	1	TGTGTTTCTCAGAAATAGCAGCAATATTATATAAAGCATGCAATCTCTATACATCG	60
Db	68685	TGTGTTTCTCTTAAATATCATGATATATTATATAAAGTATGCAATCTCTATAGAATA	68626
QY	61	CGAAGTTTAAAAAACAATAGAAATTTTACATATTTACATGGTGTGTTTATTTGGATAACA	120
Db	68625	CGAAGTATTAAACAAATATATAGAAATTTTACAAAATTAATGGTGTGTTTATTTGGATAACA	68566
QY	121	TGACAAATATTTATTTATTCATGAGTTTTTATTTGGATAGCATGACAAAATTTAATATAT	180
Db	68565	TGACAAATATTTATATATTTTCATGAGTTTTTGTGTTGATAACATGACAAAATTTAATATAT	68506
QY	181	CAGTGTATAACATGTTTGTGTTTAAATAACATGCAATTTTAA-AATCAGACATTTGTT	239
Db	68505	CAGTGTATAACATGTTTGTGTTTAAATAACATGCAATTTTAAAGAAATCATACATTTGTT	68446

QY	240	TTAAAAATCAAAATCTAATCTCTTATATACAAACGACATTCACGAAAAATTCAGGTAAAAAG	299
Db	68445	TTAAAAATCATATCTAATCTCT--TATCACAACGACAAATGATGGAGAAATTCAGGCAAAAAG	68388
QY	300	AGAAAAATAAAGATAGAGATAGAGAGATTTCTATGGAAGAAAAAGAGAGA-GAACATGTA	358
Db	68387	AGAAAAAGAGAAATGAGAGATAGAGAGATTTCTATGGAAGAAAAAGAGATAGGAACATGTA	68328
QY	359	GCTGACAAAAATAAGAGATATGATGATATATTTTATGAGAGGTGCTGAGATTTATTTTA	418
Db	68327	GCTGACAAAAATAAGAGATATGATGATATATTTATATGAGAGGTGCTGAGGA-GA-TATTTCTA	68269
QY	419	GGAGAGGGAGAGAGAAATAGAAAAAGAAAAATGACATGGTCAATCTGAA---GAAGATGAA	475
Db	68268	GGAGAGAGAGAGACAAAAGAGAAAAAGAAAAATGACATGGTGAACCTGAAGATGAAGATGAG	68209
QY	476	TTGTGTTAAAGATGAAGAGAGAAAAAGAACTCCTGCTGCTAAAGTCTCGTAAAGAAATGA	535
Db	68208	TTGTGTTCAAAGATGAAGAGAGAAAAAGAACTCCTGCTGCTAAAGTCTC-----68162	
QY	536	AAAAAGAAAAAGAGAGAAAGAGAAAGAAAGGCTAAAAATAGACTAATATTGCTCAAA	595
Db	68161	-----ATTACAAA	68153
QY	596	ATTTCTGTAGCCGACAAATACTATTTGGTCCAAGGTTATTTTGTGTATTTCTTTTGAAGTC	655
Db	68152	ATTTCTCTACAGACTAGTACTATTTGGTCCAAGGTTATTTTGTGTATTTCTTTTCAAGTC	68093
QY	656	AAAAGTATTTCTTACATATCTCTTAAAAATATAGCCGATACCAATTTTTCACACATGG	715
Db	68092	AAATGTTATTTTGTATATATCTCTATTTTCTTATATATATATATATATATATATATATAT	68033
QY	716	ACTTCTCTTATTCCTCAAAAGTCA	737
Db	68032	AGAAATTTTAAATTTTGAATTA	68011

## RESULT 13

LOCUS	AC007399	105223 bp	DNA	linear	PLN 26-APR-1999				
DEFINITION	Arabidopsis thaliana BAC F14123 from chromosome V near 69 cM, complete sequence.								
ACCESSION	AC007399								
VERSION	AC007399.1	GI:4680765							
KEYWORDS	HTG								
SOURCE	Arabidopsis thaliana (thale cress)								
ORGANISM	Arabidopsis thaliana								
REFERENCE	1 (bases 1 to 105223) Habermann, K., Nascimeto, L., Vil, M.D., Matero, A., Rodriguez, M., Shah, R., Swaby, I., Shekher, M., O'Shaughnessy, A., Huang, E.N., Spiegel, L.A., Schutz, K., Parnell, L.D., Preston, R.R., See, L.H., Dedhia, N.N. and McCombie, W.R. Arabidopsis thaliana BAC F14123 from chromosome V near 69 cM								
AUTHORS	Unpublished								
TITLE	2 (bases 1 to 105223) Habermann, K., Nascimeto, L., Vil, M.D., Matero, A., Rodriguez, M., Shah, R., Swaby, I., Shekher, M., O'Shaughnessy, A., Huang, E.N., Spiegel, L.A., Schutz, K., Parnell, L.D., Preston, R.R., See, L.H., Dedhia, N.N. and McCombie, W.R.								
JOURNAL	Direct Submission								
REFERENCE	Submitted (26-APR-1999) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724								
AUTHORS	BAC F14123 maps to near 69 cM on the Lister & Dean RI map and is assigned to YAC C1C12F8. Position 1 of F14123 is oriented toward the north telomere centromere and position 105223 is oriented toward the centromere. For more information on the mapping sequencing and annotation of F14123, please see <a href="http://www.cshl.org/arabweb/F14123-titlepage.html">http://www.cshl.org/arabweb/F14123-titlepage.html</a> . A graphic view of our annotation will also be available at this url. Gene models								
TITLE									
JOURNAL									
COMMENT									



chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em:, EMBL; Sw:, SWISSPROT; Tr:, TRMBL; Wp:, WORMPEP; Information

on the WORMPEP database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-296F5 is

from the RPCI-23 Mouse BAC Library

constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6.

#### FEATURES

source

Location/Qualifiers

1..19577

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/chromosome="4"

/clone="RP23-296F5"

/clone\_lib="RPCI-23"

#### ORIGIN

```
Query Match      10.3%; Score 99.2; DB 10; Length 19577;
Best Local Similarity 50.6%; Pred. No. 2.1e-06;
Matches 266; Conservative 0; Mismatches 258; Indels 2; Gaps 1;

QY 57 ATCGCGAAGTTTAAAAACATATAGAAATTGTTACAAATATACATGGGTTTTTATTGGAT 116
Db 12415 ACCACAACTTGGTATTGAAATACAATTATTCAACAAATAAATATGTGTATCCATGT 12474

QY 117 AACATGACAAATATTATTATTTCATGAGCTTTTATTGGATAGCATGACAAATATTAAT 176
Db 12475 AAGATGAGAAAGAACATCTGAGTCATTCCTCAAGGAATTTGGAATTCCTCAAGAGGATAA 12534

QY 177 ATATCAGTGTATAACATGTTTGTGTTCTTTAAAA--TACATGCATTTTAAAAATCAGACAT 234
Db 12535 ATACTTTCCTTAGAAAAATGATCGCGGAAAAAACCCAAAAAACCCAAAAACAAACAA 12594

QY 235 TTGTTTTTAAATCAAAATCTTAATCTCTTATATCAACGACATTTGACGCGAAAAATTCAGGTA 294
Db 12595 ACAAAGAAAAAACAACAAAAAACAACAAAAAAGAAAGAAAGAAAGAAAGAAAG 12654

QY 295 AAGAAGAAATAAAGAAATGAGAGATAGAGAGATTTCTATGGAAGAAAGAAAGAGAGACAA 354
Db 12655 AAAGGAAAGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAAAG 12714

QY 355 TGTAGGTGAACAAAAATAAGAGATATGATGATATATTTTATGAGAGTGGTGAAGATTAT 414
Db 12715 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAAAG 12774

QY 415 TTTAGGAGAGGAGAGAGAAATAGAAAAAATGACATGGTGAATCTCGAAGAAAGATGA 474
Db 12775 AGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAG 12834

QY 475 ATTGTGTTAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATG 534
Db 12835 AGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAG 12894

QY 535 AAAAAAGAAACAAAGAAAGGAAAGAAAGAAAGAAAGGCTTAAATAGA 580
Db 12895 AAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 12940
```

Search completed: November 7, 2005, 10:27:04

Job time : 4638 secs

**This Page Blank (uspto)**